EXHIBIT A
REDACTED VERSION OF DOCUMENT SOUGHT TO BE FILED UNDER SEAL

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

IN RE: FACEBOOK, INC. CONSUMER PRIVACY USER PROFILE LITIGATION, This document relates to: ALL ACTIONS	**MDL NO. 2843** CASE NO. 3:18-MD-02843-VC-JSC HON. VINCE CHHABRIA COURTROOM 4 – 17TH FLOOR SPECIAL MASTER: DANIEL GARRIE ESQ. **ORDER RE: FACEBOOK'S MOTION FOR PROTECTIVE ORDER AGAINST PRODUCTION OF API CALL LOGS** **JAMS REF. NO: 1200058674**

INTRODUCTION

1.   Pending before Special Master Garrie is Facebook's Motion for a Protective Order Against Production of API Call Logs. Facebook seeks a protective order that the data in the two tables titled ██████████████ ("Mobile Table") and ██████████████ ("Web Table") is not reasonably accessible because of undue burden and cost and that Facebook is protected from producing them.

BACKGROUND

2.   On May 6, 2020, Plaintiffs served their Third Set of Requests for Production in which they requested "all records or logs of API Calls made by Third Parties for the Content and Information of Friends of Installing Users."[1] See Exhibit A (Third Set of Requests for Production) at 10.

3.   Facebook uses APIs to allow Facebook's users to be able to access their Facebook profiles and those of their friends from local devices using Facebook's own apps (e.g., when a user pulls up their profile on a phone, the Facebook app calls Facebook's central servers through APIs to access the photos, posts, and other information displayed on that profile). See Exhibit B (Ji Declaration) at ¶ 5. Facebook also uses APIs to allow third parties to connect with its servers to send or receive information (e.g., a website can connect to Facebook's APIs to verify a user's identity rather than requiring the user to create a specific account for that site). See id. at ¶ 6.

4.   Facebook's two ██████████████ logs of API call activity are kept in two data tables: the Mobile Table, ██████████████████████████████ ██████████████████████████ ; and the Web Table, which logs API calls ████████

---

[1] API refers to "Application Programming Interface," a standard industry programming tool that allow computers and computer programs to send information to each other. Each connection to an API is referred to as a "call."

ORDER RE: FACEBOOK'S MOTION FOR PROTECTIVE ORDER AGAINST PRODUCTION OF API CALL LOGS

████████████████████████████████. See Exhibit B at ¶¶ 8-9. The Mobile Table and Web

Table reflect the number of calls made to the Facebook APIs ████████████████████████

████████████.[2] The Mobile Table and Web Table do not contain data demonstrating whether

data was returned in response to an API call, what the data that was returned may have been (if it

was returned at all), or the volume of data that may have been returned (if it was transferred at

all). Id. at ¶ 14.

     5.   Facebook preserves the Mobile Table and Web Table in "cold storage" in a highly

compressed, non-readable format. Id. at ¶ 17. In cold storage, the data in the tables cannot be

reviewed or analyzed. Id. at ¶ 18. Facebook maintains the Mobile Table from ████████████

████████, and it contains approximately ████████████ of data in highly compressed form.[3] See

Exhibit C (Motion for Protective Order Against Production of API Call Logs) at 5. Facebook

maintains the Web Table from ████████████████████, and it contains approximately ████

████████ of data in highly compressed form. Id.

     6.   To analyze the data in the Mobile Table and Web Table, Facebook data scientists

must first restore a partition of the data to "warm storage"—a more accessible form of data

storage that makes the data readable. Id. at ¶¶ 20. Due to limitations on computing and server

capacity, only a limited number of partitions can be restored and analyzed at any given time. Id.

It takes about one day of processing to restore ████████████ of data from the Tables to warm

storage. See id. at ¶ 18.

---

[2] The data fields included in the Mobile Table include ████████████████████████████████
████████████████████████████. See Exhibit A to Ji Declaration. The data fields in the Web Table include
██
████████████████████. See Exhibit C to Ji Declaration.

[3] A petabyte is one million gigabytes.

**ORDER RE: FACEBOOK'S MOTION FOR PROTECTIVE ORDER AGAINST PRODUCTION OF API
CALL LOGS**

7.   Facebook data scientist Mengge Ji estimated it would take approximately four

years of dedicated full-time work for her to restore one partition, run searches across that

restored partition, and then repeat ███████████████████████████████████, not

including the additional time it would take to export data for production. Id. at ¶ 35.

8.   Facebook also maintains a smaller table called ██████████████████ ("Method

Table") that provides summary data █████████████ of API calls ██████████████

████████████████████████████.[4] Id. at ¶ 53.

9.   On October 6, 2021, Special Master Daniel Garrie ("Special Master Garrie") and

Hon. Gail Andler (Ret.) declared impasse on the issue of whether Facebook is required to

produce all API call log data.

10.   On October 18, 2021, Facebook submitted its opening brief to Special Master

Garrie on this issue. See Exhibit C. Facebook argues that (a) the Mobile Table and Web Table are

not reasonably accessible because of undue burden and cost; and (b) Plaintiffs cannot show good

cause for the production of the inaccessible Mobile Table and Web Table data because the Tables

are outside the scope of discovery permitted by Rule 26(b)(1) and Facebook is willing to produce

alternative sources of information about the content of API call logs. Id.

11.   On October 28, 2021, Plaintiffs submitted their Opposition to Facebook's Motion

for a Protective Order to Stop Preserving Certain API Call Logs. See Exhibit D. Plaintiffs argue,

among other things, that (a) Facebook fails to demonstrate good cause for entry of a protective

order; and (b) Facebook fails to provide the information necessary to ascertain the benefits of

---

[4] The data fields included in the Method Table are ████████████████████████████

████████████████████████████. See Exhibit E to Ji Declaration.

**ORDER RE: FACEBOOK'S MOTION FOR PROTECTIVE ORDER AGAINST PRODUCTION OF API
CALL LOGS**

preservation. Id. Plaintiffs also propose that Facebook provide additional information regarding the Mobile Table and Web Table to facilitate a targeted sampling of data from the Tables.

12.   On November 2, 2021, Facebook submitted its Reply in Support of Motion for a Protective Order Against Production of API Call Logs. See Exhibit E. Facebook argues, among other things, that (a) Plaintiffs fail to show that the relevance, importance, or benefit of data from the Mobile and Web Tables outweighs the burden of production; (b) Plaintiffs' proposed solution of sampling the Tables is unduly burdensome when balanced against the appropriate proportionality considerations and would yield only irrelevant data; and (c) Plaintiffs have not shown that the Method Table is an inadequate alternative to the Mobile and Web Tables.

FINDINGS

13.   According to Rule 26(b)(2)(B):

> A party need not provide discovery of electronically stored information from sources that the party identifies as not reasonably accessible because of undue burden or cost. On motion to compel discovery or for a protective order, the party from whom discovery is sought must show that the information is not reasonably accessible because of undue burden or cost. If that showing is made, the court may nonetheless order discovery from such sources if the requesting party shows good cause, considering the limitations of Rule 26(b)(2)(C).

14.   "In general, [ESI is] inaccessible [if it] 'is not readily useable and must be restored to an accessible state before the data is usable.'" U.S. ex rel. Carter v. Bridgepoint Educ., Inc., 305 F.R.D. 225, 240 (S.D. Cal. 2015) (quoting Zubulake v. UBS Warburg LLC, 217 F.R.D. 309, 320 (S.D.N.Y. 2003)).

15.   Special Master Garrie finds that the data in the Mobile Table and Web Table is not reasonably accessible because it is not readily usable in its "cold storage" state and must be restored to "warm storage" in order to be searched and analyzed (i.e. usable). See Exhibit B at ¶¶ 17-20. Special Master Garrie finds that the time and cost of restoring, analyzing, and producing the data

4

in the Mobile Table and Web Table would place an undue burden on Facebook when balanced against the appropriate proportionality considerations.

16.   Moreover, Special Master Garrie finds that restoring the data in the Mobile Table and Web Table for the purpose of sampling, as Plaintiffs propose, would also place an undue burden on Facebook, balanced against the appropriate proportionality considerations, as it would require restoring the majority of partitions to warm storage, searching the partitions individually, and exporting a large volume of data. Id. at ¶ 19, 28, 35. This process would most likely take years to complete.

17.   Special Master Garrie finds that Plaintiffs have not established good cause for producing data from the Mobile Table and Web Table because comparable data is available through the accessible Methods Table. Id. at ¶ 53. Moreover, Special Master Garrie finds that the Mobile Table and Web Table are not likely to yield relevant information regarding the sharing of user data with third parties because the Mobile Table and Web Table do not contain data demonstrating whether data was returned in response to an API call, what the data that was returned may have been (if it was returned at all), or the volume of data that may have been returned (if it was transferred at all). Id. at ¶ 14.

//

//

//

//

//

//

ORDER RE: FACEBOOK'S MOTION FOR PROTECTIVE ORDER AGAINST PRODUCTION OF API CALL LOGS

ORDER

18.   Facebook's motion for a protective order against producing API call log data from the Mobile Table and Web Table is granted.

19.   No later than December 1, 2021, Facebook is to produce the data in the Methods Table.

IT IS SO ORDERED.

Monday, November 8, 2021

Daniel Garrie
Discovery Special Master

ORDER RE: FACEBOOK'S MOTION FOR PROTECTIVE ORDER AGAINST PRODUCTION OF API CALL LOGS

EXHIBIT A

Lesley E. Weaver (SBN 191305)
BLEICHMAR FONTI & AULD LLP
555 12th Street, Suite 1600
Oakland, CA 94607
Tel.: (415) 445-4003
Fax: (415) 445-4020
lweaver@bfalaw.com

Derek W. Loeser (admitted *pro hac vice*)
KELLER ROHRBACK L.L.P.
1201 Third Avenue, Suite 3200
Seattle, WA 98101
Tel.: (206) 623-1900
Fax: (206) 623-3384
dloeser@kellerrohrback.com

Plaintiffs' Co-Lead Counsel

Additional counsel listed on signature page

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA

IN RE: FACEBOOK, INC. CONSUMER PRIVACY USER PROFILE LITIGATION	MDL No. 2843 Case No. 18-md-02843-VC
This document relates to: ALL ACTIONS	**PLAINTIFFS' THIRD SET OF REQUESTS FOR PRODUCTION TO DEFENDANT FACEBOOK, INC.** Judge: Hon. Vince Chhabria Courtroom: 4, 17th Floor

PROPOUNDING PARTY: Plaintiffs

RESPONDING PARTY: Facebook

SET NUMBER: Third (3)

Plaintiffs hereby propound the following requests for production of documents to Defendant Facebook, Inc. ("Facebook"), pursuant to Federal Rules of Civil Procedure 26 and 34, and request that Facebook produce the documents and electronically-stored information set forth herein within thirty (30) days of service of these requests, at Bleichmar Fonti & Auld LLP, 555 12th Street, Suite 1600, Oakland, CA 94607.

INSTRUCTIONS

1.   You shall respond to these requests for the production of documents in a manner consistent with the Federal Rules of Civil Procedure and the following instructions:

2.   In responding to each document request, furnish all responsive documents available at the time of production, including documents in your possession, custody or control, and in the possession, custody or control of your agents, employees, partners, representatives, subsidiaries, affiliates, investigators, or by your attorneys or their agents, employees or investigators.

3.   If any otherwise responsive document was, but is no longer, in existence or in your possession, custody or control, identify the type of information contained in the document, its current or last known custodian, the location/address of such document, the identity of all persons having knowledge or who had knowledge of the document and describe in full the circumstances surrounding its disposition from your possession or control.

4.   This is a continuing request for the production of documents and requires supplemental responses as provided for in the Federal Rules of Civil Procedure. If, after making your initial production, you (or any other persons acting on your behalf) obtain or become aware

of any further documents responsive to any document request, you are required to produce such additional documents to plaintiffs. Each supplemental response shall be served on plaintiffs no later than thirty days after the discovery of the further information.

5.   You shall produce the original of each document described below or, if the original is not in your custody, then a copy thereof, and in any event, all non-identical copies which differ from the original or from the other copies produced for any reason, including, without limitation, the making of notes thereon.

6.   Documents shall be produced as kept in the regular course of business together with the original folders, binders, boxes or other containers in which they were maintained.

7.   All documents or things that respond in whole or in part to any portion of these requests are to be produced in their entirety, including attachments and their enclosures.

8.   Documents attached to each other should not be separated.

9.   Documents not otherwise responsive to any particular document request shall be produced if such documents mention, discuss, refer to, or explain the documents called for by any document request, or if such documents are attached to documents called for by any document request.

10.   Documents shall be produced in such fashion as to identify the custodian of each document.

11.   Identify the source of each document produced, by identifying: (a) all of the person(s) who possessed the document; (b) the positions or titles of any such individuals; and (c) all of the divisions and departments where each document was located. If you are unable to determine the individual(s) who possessed the document, identify the department and division where the document was located when produced.

12.   If you claim any form of privilege, whether based on statute or otherwise, as a ground for not producing any document, state the following:

   a. The date of the document;

   b. The name, the present or last known home and business address, the telephone numbers, the title (or position), and the occupation of those individuals who prepared, produced, reproduced or who were recipients of said document;

   c. A description of the document sufficient to identify it without revealing the information for which the privilege is claimed;

   d. The nature of the privilege asserted;

   e. The factual basis upon which you claim any such privilege;

   f. The location of the document; and

   g. The custodian of the document.

13.   To the extent you object to any document request, you must provide specific responses as to what portion of the request you object to and state expressly why you will not respond to such request in sufficient detail to permit the Court to determine the validity of the objection. Responsive documents to which your objection does not apply should be produced.

14.   If you claim that all or any part of any document request, the Definitions, or Instructions is vague or ambiguous, please identify the specific language you consider vague or ambiguous and state the interpretation of the language in question you used to frame your response.

15.   Each document requested herein is to be produced in its entirety and without deletion or excision, regardless of whether you consider the entire document to be relevant or responsive to any document request. If you have removed, excised or deleted any portion of a

document, stamp the word "REDACTED" on each page of the document that you have redacted.

Redactions should be included on the privilege log described in Instruction No. 13, above.

16.   One copy of each document should be produced. A document that varies in any

way from the original or from any other copy, including drafts or a document with handwritten

notations or deletions constitutes a separate document and must be produced, whether or not the

original is in your possession, custody or control. Color (i.e., not black and white) originals

should be produced in color. If any identical copy cannot be produced for any reason (e.g., faint

writing, erasures, etc.), produce the original.

17.   Indicate the origin of each document and number each document with

consecutive Bates numbers.

DEFINITIONS

Unless otherwise stated, the terms set forth below are defined as follows and shall be

used in construing the meaning of these requests for the production of documents.

1.   The use of the singular shall be deemed to include the plural, and the use of one

gender shall include all others, as appropriate, in the context.

2.   The present tense of a verb includes its past tense, and vice versa.

3.   "And" and "or" are to be construed conjunctively and disjunctively, as necessary,

to bring within the scope of this request for production all responses that might otherwise be

construed to be outside its scope.

4.   "Any" and "all" mean each and every.

5.   "App" means an interactive software application developed to utilize the core

technologies of the Facebook social networking platform.

6.   "Business Partners" refers to the third parties with whom Facebook partnered to

develop and integrate Facebook on a variety of devices and operating systems, including, but not

limited to, the third parties listed in paragraph 484 of Plaintiffs' First Amended Consolidated Complaint.

7.   "Call" means an automated call or other data-retrieval request delivered to or through the Facebook Platform, including through the Graph.

8.   "Content and Information" refers to the definition in footnote 2 of the FAC, referring to "content" and "information" as Facebook's Statements of Rights and Responsibilities have defined those terms. In brief, Facebook has generally used "information" to mean facts and other information about Facebook Users, including the actions they take, and "content" to mean anything Facebook Users post on Facebook that would not be included in the definition of "information." Content and Information also includes both personally identifiable content and information and anonymized content and information that is capable of being de-anonymized. *See* FAC ¶¶ 223-224. Content and Information includes data that identifies, relates to, describes, is capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular Facebook User, including:

    a. Identifiers such as a real name, alias, postal address, unique personal identifier, online identifier, Internet Protocol address, email address, account name, social security number, driver's license number, passport number, or other similar identifiers.

    b. Characteristics of protected classifications under California or federal law.

    c. Commercial information, including records of personal property, products or services purchased, obtained, or considered, or other purchasing or consuming histories or tendencies.

    d. Biometric information.

    e. Internet or other electronic network activity information, including, but not

limited to, browsing history, search history, and information regarding a

consumer's interaction with an Internet Web site, application, or

advertisement.

f.   Geolocation data.

g.   Audio, electronic, visual, thermal, olfactory, or similar information.

h.   Professional or employment-related information.

i.   Education information, defined as information that is not publicly available

personally identifiable information as defined in the Family Educational

Rights and Privacy Act (20 U.S.C. section 1232g, 34 C.F.R. Part 99).

j.   Inferences drawn from any of the information identified in this paragraph to

create a profile, dossier, or similar collection of information about a

consumer reflecting the consumer's preferences, characteristics,

psychological trends, predispositions, behavior, attitudes, intelligence,

abilities, and aptitudes.

9.   "Data Warehouse" refers to any warehousing framework developed by the Data

Infrastructure Team at Facebook and may also be known as the "Hive."

10.   "Database" refers to any organized collection of information that is stored

electronically.

11.   "Document" or "Documents" is defined to include any document, ESI, or

Electronic Media stored in any medium, and is synonymous in meaning and equal in scope to

the usage of this term in Federal Rule of Civil Procedure 34(a)(1)(A), including, but not limited

to programming source code, electronic or computerized data compilations, Communications,

electronic chats, instant messaging, documents created through Workplace by Facebook,

encrypted or self-destructing messages, messages sent via Facebook messenger, email

Communications, other electronically stored information from Personal computers, sound

recordings, photographs, and hard copy Documents maintained in your Personal files.

12.　　"Facebook," "Defendant," "You," or "Your" shall mean Facebook, Inc. and any

of its executives, directors, officers, employees, partners, members, representatives, agents

(including attorneys, accountants, consultants, investment advisors or bankers), and any other

Person purporting to act on its behalf. In the case of business entities, these defined terms

include parents, subsidiaries, affiliates, predecessor entities, successor entities, these defined

terms include parents, subsidiaries, affiliates, predecessor entities, successor entities, divisions,

departments, groups, acquired entities and/or related entities or any other entity acting or

purporting to act on its behalf.

13.　　"Facebook Platform" means a set of APIs and services provided by Facebook that

enable websites and applications to (a) retrieve Content and Information relating to Facebook

Users made available by Facebook, (b) retrieve authorized Content and Information relating to

Facebook Users from other applications or websites, and (c) provide Content and Information

relating to Facebook Users to Facebook.

14.　　"Facebook User" means persons who maintain a Facebook account.

15.　　"Friends of Installing User" refers to Facebook Users who did not install a

particular App, but whose Content and Information became accessible to that App because they

were Facebook friends with an Installing User.

16.　　"Hive" refers to the open source, peta-byte scale data warehousing framework

based on Hadoop that was developed by the Data Infrastructure Team at Facebook, also known

as the Data Warehouse.

17.   "Including" means "including but not limited to," or "including, without limitation." Any examples which follow these phrases are set forth to clarify a request, definition, or instruction but not to limit it.

18.   "Installing User" refers to the Facebook User who installed a particular App via his or her Facebook account.

19.   "Platform" refers to the services, tools, and products provided by Facebook to third parties to create their own applications and services that access data in Facebook.

20.   "Privacy Controls" means the audience selectors that control what information in a Facebook User's profile can be viewed by other Facebook Users, and includes Profile Privacy Settings, Profile Privacy Controls, Publisher Privacy Controls, and the like.

21.   "Privacy Cross-Functional Team" means the team of employees working on Facebook's Privacy Program to regularly assess risks and controls on an ongoing basis of the Privacy Program.

22.   "Privacy Governance Team" means the designated team of employees directly responsible for the Facebook Privacy Program.

23.   "Privacy Program" means the program established, implemented, and maintained by Facebook pursuant to the 2012 Consent Order between Facebook and the Federal Trade Commission.

24.   "Relating to," "relate to," "referring to," "refer to," "reflecting," "reflect," "concerning," or "concern" means all Documents which comprise, explicitly or implicitly refer to, were reviewed in conjunction with, or were created, generated or maintained as a result of the subject matter of the request, including, but not limited to, all Documents which reflect, record, memorialize, embody, discuss, evaluate, consider, review or report on the subject matter of the request.

25.   "Third Parties" include the following:

    a.   Apps, App Developers, Whitelisted Apps, and Business Partners, as those

        terms are used in the FAC;

    b.   Any person that develops an application, software experience, game, or

        website that accesses Content and Information from Facebook's API or other

        Facebook software; and

    c.   Any person with which Facebook has or had an integration partnership.

26.   "Version" refers to both major version and minor version of an API (*e.g.*, Version

2.0 is different from Version 2.1).

27.   Capitalized terms and acronyms not specifically defined herein have the same

definition as in the FAC.

RELEVANT TIME PERIOD

The relevant time period for each Document Request is January 1, 2007 through the

present (the "Relevant Time Period"), unless otherwise specifically indicated, or portion thereof.

Each Document Request shall be interpreted to include all documents and information that relate

to the Relevant Time Period or otherwise-specified period, even if such documents or

information was prepared or published outside of the Relevant Time Period or otherwise-

specified period. If a document prepared before or after this period is necessary for a correct or

complete understanding of any document covered by a request, you must produce the earlier or

subsequent document as well. If any document is undated and the date of its preparation cannot

be determined, the document shall be produced if otherwise responsive to the production request.

DOCUMENT REQUESTS

REQUEST FOR PRODUCTION NO. 37

All records or logs of API Calls made by Third Parties for the Content and Information of Friends of Installing Users, including:

    a. For each Version of each API, Calls for Facebook Users' and Friends' Content and Information;

    b. For each Call, information about the endpoint that originated the Call, including the endpoint's IP (Internet Protocol) address and whether the endpoint is a user device or a Third Party's server;

    c. The response status of those Calls (*e.g.* success or reason for failure);

    d. The Content and Information delivered to Third Parties on each API in response to these Calls; and

    e. For each Version of each API, the format in which such Content and Information was delivered to Third Parties and whether it was stored.

REQUEST FOR PRODUCTION NO. 38

All Documents concerning privacy settings, sharing controls, and deletion capabilities (including permanent deletion) available to any Facebook employee, officer, or director using the Platform that were not available to all Facebook Users.

REQUEST FOR PRODUCTION NO. 39

For each Database containing information about Facebook Users—including, but not limited to, those based on the Hive and/or any Data Warehouse—all manuals, schematics, and other Documents describing the content, structure, or organization of each such Database.

REQUEST FOR PRODUCTION NO. 40

Documents sufficient to identify all Databases or Data Warehouses which contain information about Facebook Users.

REQUEST FOR PRODUCTION NO. 41

All Documents relating to the deletion, loss, destruction or spoliation of any Documents or data responsive to any request for production propounded in this action.

REQUEST FOR PRODUCTION NO. 42

All Documents You relied upon in answering any interrogatory Plaintiffs propound in this action.

REQUEST FOR PRODUCTION NO. 43

All privilege logs, interrogatory responses, written reports, correspondence and deposition transcripts from any formal or informal inquiry or investigation by a governmental entity or regulator in the United States or United Kingdom relating to whether Facebook Users' Content and Information was accessed or obtained by any Third Parties without proper consent or authorization, including but not limited to all inquiries or investigations arising out of the Cambridge Analytica Scandal, the FTC Consent Order, and any inquiry or investigation related to the settlement agreement with the FTC announced on July 24, 2019.

REQUEST FOR PRODUCTION NO. 44

Documents sufficient to show all payments You received from all Third Party Apps or App developers who had access to the Content and Information of Friends of Installing Users in the United States or United Kingdom.

REQUEST FOR PRODUCTION NO. 45

Documents sufficient to show the value of any data You received from all Third Party Apps or App developers who had access to the Content and Information of Friends of Installing Users in the United States or United Kingdom, including any estimates of value that You made.

REQUEST FOR PRODUCTION NO. 46

Documents sufficient to show all costs associated with Facebook granting Third Party Apps or App developers access to the Content and Information of Friends of Installing Users in the United States or United Kingdom.

REQUEST FOR PRODUCTION NO. 47

Documents sufficient to show all payments You received from all Business Partners who Facebook granted access to the Content and Information of Facebook Users in the United States or United Kingdom.

REQUEST FOR PRODUCTION NO. 48

Documents sufficient to show the value of any data You received from all Business Partners who Facebook granted access to the Content and Information of Facebook Users in the United States or United Kingdom, including any estimates of value that You made.

REQUEST FOR PRODUCTION NO. 49

Documents sufficient to show all costs associated with Facebook granting Business Partners access to the Content and Information of Facebook Users in the United States or United Kingdom.

REQUEST FOR PRODUCTION NO. 50

All Documents regarding Your assessment of the financial impact of changes to user data practices and product changes that restrict developer access to certain user data identified in Your 2018 Form 10-K at pages 9 and 24.

REQUEST FOR PRODUCTION NO. 51

Documents sufficient to show the costs You incurred to enforce policies regarding Third Party access to and use of Your Facebook Users' Content and Information.

REQUEST FOR PRODUCTION NO. 52

All Documents regarding all monitoring controls and audit processes in operation to detect misuse of the Facebook Platform throughout the Class Period.

REQUEST FOR PRODUCTION NO. 53

All Documents regarding privacy decisions made by the Privacy Governance Team and the Privacy Cross-Functional Team related to Facebook's Privacy Program.

REQUEST FOR PRODUCTION NO. 54

All Documents regarding all enhanced privacy features established, implemented, and maintained by Facebook and how Facebook established, implemented, and maintained such features.

REQUEST FOR PRODUCTION NO. 55

All Documents regarding all steps taken by Facebook to establish, implement, and maintain controls related to how Third Parties accessed Content and Information, including the approval process Facebook utilized to grant such access.

REQUEST FOR PRODUCTION NO. 56

All Documents regarding Facebook's ongoing monitoring of its Privacy Program, including how Privacy Program controls were reviewed and updated, what methodologies Facebook utilized to test these controls, and how Facebook designed and implemented intake, detection, handling, response, remediation, and reporting of all privacy incidents such as misuse of user data by Third Parties.

Dated: May 6, 2020

Respectfully submitted,

KELLER ROHRBACK L.L.P.

BLEICHMAR FONTI & AULD LLP

By: */s/ Derek W. Loeser*
       Derek W. Loeser

By: */s/ Lesley E. Weaver*
       Lesley E. Weaver

Derek W. Loeser (admitted *pro hac vice*)
Lynn Lincoln Sarko (admitted *pro hac vice*)
Gretchen Freeman Cappio (admitted *pro hac vice*)
Cari Campen Laufenberg (admitted *pro hac vice*)
David Ko (admitted *pro hac vice*)
Adele A. Daniel (admitted *pro hac vice*)
Benjamin Gould (SBN 250630)
1201 Third Avenue, Suite 3200
Seattle, WA 98101
Tel.: (206) 623-1900
Fax: (206) 623-3384
dloeser@kellerrohrback.com
lsarko@kellerrohrback.com
gcappio@kellerrohrback.com
claufenberg@kellerrohrback.com
bgould@kellerrohrback.com
dko@kellerrohrback.com
adaniel@kellerrohrback.com

Lesley E. Weaver (SBN 191305)
Anne K. Davis (SBN 267909)
Joshua D. Samra (SBN 313050)
Matthew P. Montgomery (SBN 180196)
Angelica M. Ornelas (SBN 285929)
555 12th Street, Suite 1600
Oakland, CA 94607
Tel.: (415) 445-4003
Fax: (415) 445-4020
lweaver@bfalaw.com
adavis@bfalaw.com
jsamra@bfalaw.com
mmontgomery@bfalaw.com
aornelas@bfalaw.com

Christopher Springer (SBN 291180)
801 Garden Street, Suite 301
Santa Barbara, CA 93101
Tel.: (805) 456-1496
Fax: (805) 456-1497
cspringer@kellerrohrback.com

Plaintiffs' Co-Lead Counsel

EXHIBIT B

GIBSON, DUNN & CRUTCHER LLP
Orin Snyder (*pro hac vice*)
  osnyder@gibsondunn.com
200 Park Avenue
New York, NY 10166-0193
Telephone: 212.351.4000
Facsimile: 212.351.4035

Kristin A. Linsley (SBN 154148)
  klinsley@gibsondunn.com
Martie Kutscher (SBN 302650)
  mkutscherclark@gibsondunn.com
555 Mission Street, Suite 3000
San Francisco, CA 94105-0921
Telephone: 415.393.8200
Facsimile: 415.393.8306

GIBSON, DUNN & CRUTCHER LLP
Deborah Stein (SBN 224570)
  dstein@gibsondunn.com
333 South Grand Avenue
Los Angeles, CA 90071-3197
Telephone: 213.229.7000
Facsimile: 213.229.7520

Joshua S. Lipshutz (SBN 242557)
  jlipshutz@gibsondunn.com
1050 Connecticut Avenue, N.W.
Washington, DC 20036-5306
Telephone: 202.955.8500
Facsimile: 202.467.0539

Attorneys for Defendant Facebook, Inc.

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

IN RE: FACEBOOK, INC. CONSUMER PRIVACY USER PROFILE LITIGATION, This document relates to: ALL ACTIONS	CASE NO. 3:18-MD-02843-VC **DECLARATION OF MENGGE JI IN SUPPORT OF FACEBOOK, INC.'S MOTION FOR A PROTECTIVE ORDER AGAINST PRODUCTION OF API CALL LOGS** Discovery Special Master Daniel Garrie, Esq.

DECLARATION OF MENGGE JI IN SUPPORT OF
FACEBOOK, INC.'S MOTION FOR A PROTECTIVE ORDER AGAINST
PRODUCTION OF API CALL LOGS

HIGHLY CONFIDENTIAL

I, Mengge Ji, declare as follows:

1.   I am a Data Scientist at Facebook, Inc. ("Facebook"). My job responsibilities include, among other things, understanding and working with Facebook's data systems, writing queries and conducting analyses of these data, researching Facebook's data and related technologies, and locating, analyzing, and exporting data for production in litigation and other legal matters. I submit this declaration in support of Facebook's motion for a protective order as against the production of certain of Facebook's API call logs. Unless otherwise stated, I have personal knowledge of the facts set forth herein, and, if called as a witness, I could and would competently testify thereto.

2.   My understanding is that Plaintiffs in this matter have requested that Facebook produce to them all "logs of API Calls made by Third Parties for the Content and Information of Friends of Installing Users," including specific information about, among other things, "the response status of those Calls (*e.g.*, success or reason for failure)" and "[t]he Content and Information delivered to Third Parties on each API in response to these Calls."

3.   I also understand that Plaintiffs have issued additional requests for information relating to the number of API calls made by certain "Third Parties" and "Business Partner[s]," including "the number of calls . . . made to each such API . . . each month" and "the volume of data transferred from each such API . . . each month."

1.   Background on Facebook's APIs and Call Logs

4.   Application Program Interfaces ("APIs") are standard computing protocols used throughout the digital world. Each time a user visits an app, the user necessarily interacts through an API. And every company that allows third parties to communicate with their servers

2   **HIGHLY CONFIDENTIAL**

uses APIs. As just one example, email programs like Outlook or Mail on iOS communicate with

the user's email provider (Yahoo, Gmail, etc.) through APIs to obtain the user's messages.

5.   Facebook has a series of APIs that allow for data to be transferred to and from

Facebook's Platform. The Facebook Platform is powered by a series of databases that Facebook

relies upon to host and run the Facebook products and which work in tandem to provide

Facebook users a seamless experience as they access different categories and types of content

posted on the Platform, by themselves or other users. APIs facilitate the transfer of data from the

databases that support the Platform to websites and applications, including Facebook's own,

proprietary mobile apps. For instance, if a Facebook user pulls up her profile photo via the

Facebook mobile app, the Facebook mobile app will "call" a Facebook API to request that photo

to be displayed locally on the user's phone.

6.   Third party mobile apps and websites can also connect to Facebook through its

APIs. For instance, if a user chooses to log in to NYTimes.com with his Facebook account

instead of creating an account for that site, NYTimes.com will call the Facebook APIs to verify

the user's identity. Facebook has a series of public APIs that are published on its developer

website (https://developers.facebook.com/), and app developers can use them or request to use

them to send or receive information from the Facebook Platform.

7.   Facebook maintains internal logs that contain the number of API calls made to the

Facebook Platform. Two of the data tables that contain ▮▮▮▮▮▮▮ information Facebook

maintains regarding API call activity are the ▮▮▮▮▮▮ table and the ▮▮▮▮▮▮ table.

2.   The ▮▮▮▮▮▮ and ▮▮▮▮▮▮ Tables

8.   The ▮▮▮▮▮▮ table contains logs of API calls ▮▮▮▮▮▮▮▮▮

▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮

HIGHLY CONFIDENTIAL

██

██ The data in

the ███████████ table dates back to ██████████. As of October 6, 2021, Facebook had

███████████ of data in this table.

9.   The categories of data logged in the ███████████ table include, among other

things, ██

██ A true and

correct list of the data fields reflected in the ███████████ table is attached hereto as **Exhibit

A**. A true and correct sample of ███████████ from the ███████████ table, with █████

███████████████ redacted, is attached hereto as **Exhibit B**.

10.   The ███████████ table contains logs of API calls ███████████████████

███

██. The

data in the ███████████ table dates back to ██████████. As of October 6, 2021, Facebook

had ███████████ of data in this table.

11.   The categories of data logged in the ███████████ table include, among other

things, ██

████████████████████████████████. A true and correct list of the data fields

reflected in the ███████████ table is attached hereto as **Exhibit C**. A true and correct sample

of ten lines of data from the ███████████ table, with the ███████████████████ redacted,

is attached hereto as **Exhibit D**.

12.   Several of the fields in each of these tables do not contain ███████████████

███

HIGHLY CONFIDENTIAL

███████████████████████. This is an extra step of analysis that would need to be conducted if these tables were ever to be used or analyzed outside of Facebook's environment.

13.   The ████████ and ████████ tables do not contain data demonstrating whether data was returned in response to an API call, what the data that was returned may have been (if it was returned at all), or the volume of data that may have been returned (if it was transferred at all).

14.   I understand that there are many reasons why data would not be returned in response to an API call. These reasons include, but are not limited to, ████████████ ██████████████████████████████████████ █████████████████████████████████████ ████████████████████████████████████

A.   Storage of the ████████ **and** ████████ **tables**

15.   The ████████ and ████████ tables are stored in a highly compressed state in Facebook's internal decentralized, structured data warehouse, known as "Hive." Hive is designed to ingest and process data for internal analysis, not to export data into files that can be transferred outside of Facebook. As discussed further below, accessing and analyzing large data tables in Hive is a multi-step process that can take a considerable amount of time, depending on the volume of data being accessed and the complexity of the analysis or querying needed.

16.   The active ████████ and ████████ tables are currently maintained pursuant to ███ business retention periods █████████████████████ ████████ ██████████████████████████████████ ████████████████████████████████████

HIGHLY CONFIDENTIAL

███

██

17.   ██

██

█████████████████████████████████████

18.   Tables in cold storage cannot be reviewed or analyzed while they remain in cold

storage. If Facebook wants to conduct any analysis of data in cold storage, a Facebook data

scientist must first restore it from cold storage to an accessible location, referred to as "warm"

storage. It takes approximately 1 day of processing to restore ████████████ of cold storage

data. Restored cold storage data is only made available to legal team members.

19.   The ████████ and ████████ tables are stored in "partitions" or

segments. These tables are partitioned ███████████████████████████████████████

████████   Because each ██ is partitioned off from the others, Facebook cannot quickly or

efficiently run searches across multiple ███████████████████. Instead, Facebook's

system is optimized for running searches within an individual partition. As a result, Facebook

cannot search these entire tables for a specific value, for instance a ████████████, while in

cold storage, and instead must restore individual partitions to warm storage and then search the

partitions individually. Due to the same limitations on accessing and searching multiple

partitions of data, Facebook also cannot filter these tables for ███████████████████

████████████████████████ without first restoring partitions from cold

storage piecemeal.

20.   Accordingly, to search or analyze the data in either the ████████████ table or

████████ tables, a Facebook data scientist must first restore a partition of the data to warm

6   **HIGHLY CONFIDENTIAL**

storage. Due to limitations on computing and server capacity, only a limited number of partitions can be restored and analyzed at any given time.

21.   The ████████ and ████████ tables are very large data tables, even in comparison to other data sets at Facebook and even when compressed in Facebook's specialized data storage environment. The volume of the ████████ table is currently ████ in its most highly-compressed form in cold storage. The volume of the ████████ table is currently ████ in its most highly-compressed form. Maintaining the high volume of this data ████████ is beginning to tax Facebook's data storage system, which is one of the largest and most sophisticated data storage systems in the world. For instance, I understand that it is costing Facebook approximately ████████ to store the ████████ data in its compressed form.

B.   Sampling of the ████████ table

22.   In order to assess how long it might take to produce data from the ████████ table to the Plaintiffs in this action, I recently restored and analyzed one partition of data from the ████████ table dated ████████

23.   The volume of the ████████ partition is ████████ in cold storage. For reference, 1 TB is four-times the size of a standard laptop computer hard drive of 256 GB, so the *compressed* volume of the data for this one day equals ████████. The size of the data in each partition in the ████████ table varies ████████ ████████. For example, the most recent partitions of the table in cold storage are around ████ each.

24.   In order to analyze the ████████ data, I first restored this partition to warm storage. The length of time it takes to restore data from cold storage depends on how many

HIGHLY CONFIDENTIAL

requests are pending at the same time and the computing power available, which varies

depending on the time of day. In this instance, it took approximately 2 hours of processing time

to restore this single partition into warm storage. Based on the volume of this partition and

others in this table, I estimate that Facebook presently has the computing and server capacity to

host up to five partitions of this table in warm storage at a time. As a result, Facebook can

analyze no more than five partitions of this table at a given time in the ordinary course. In other

words, an effort to simultaneously analyze more than five partitions of this data would displace

and disrupt all other efforts by the legal team to perform other data analyses for this and other

matters, including for the ▮▮▮▮▮▮ table.

     25.   The ▮▮▮▮▮ partition of the ▮▮▮▮▮▮ table contains

▮▮▮▮▮▮▮▮▮▮▮. As a point of reference, a Microsoft Excel file can only

hold up to 1 million rows of data. If translated into Excel, then, this one partition of the

▮▮▮▮▮ table would comprise more than ▮▮▮▮ Excel files.

     26.   As I mentioned in paragraph 8 above, this table only contains logging for API

calls ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮

▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮

The partition only contained data for ▮▮▮▮▮. It took 1 hour and 14 minutes for this

search to complete.

     27.   ▮▮▮▮▮▮▮▮▮▮▮▮▮▮

▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮

▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮

▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮

▮▮▮▮

                                                                         HIGHLY CONFIDENTIAL

28.   Based on the results of this exercise, I estimate it would take 2,378 hours (█████ █████████ in cold storage) to restore each of the partitions in the ███████████ table to warm storage and additional 1,279 hours (████ █████████) to run a single search on each. Assuming I worked exclusively on this task for 8-hours per day, 260 days each year, it would take nearly two years for me to run this same search on ██████████ of this data (through the ██████████ partition). This is also a conservative estimate, ████████████████████████ ██. This estimate also does not account for the other demands on Facebook's personnel and resources (including the need to use these same systems to respond to requests for other legal matters), occasional outages, or many other variables.

C.   Sampling of the ████████████ **table**

29.   I also restored and recently analyzed one partition of data from the ██████████ table dated █████████████

30.   The volume of the ███████████ partition is █████ in cold storage. The size of the data in each partition in the ██████████ table varies ████████████████████████ ███ For example, the most recent partitions of the table in cold storage are around ████ each.

31.   In order to analyze the ███████████ data, I first restored this partition to warm storage. The length of time it takes to restore data from cold storage depends on how many requests are pending at the same time and the computing power available, which varies depending on the time of day. In this instance, it took approximately ██ hours of processing time to restore this partition into warm storage for analysis. Based on the volume of this partition and others in this table, I estimate that Facebook presently has the computing and server

9   **HIGHLY CONFIDENTIAL**

capacity to host 20 partitions of this table in warm storage at a time. As a result, Facebook can only analyze 20 partitions of this table at a given time in the ordinary course.

32.   The ██████████ partition of the ████████ table contains ████████████ rows of information. If exported to Excel, then, this partition of the ██████████ table would alone comprise █████ Excel files.

33.   I ran a search for ██ ████████████████████████████████ The partition only contained data for ██████████████. It took 43 minutes for this search to run.

34.   Of the ████████████ of data in this table, ████████████ of data were returned for these ██ ████████

35.   Based on the results of this exercise, I estimate it would take 2,970 hours (2.5 hours per partition in cold storage) to restore each of the partitions in the ██████████ table to warm storage and an additional 935 hours (45 minutes per partition) to run a single search on each. Assuming I worked exclusively on this task for 8-hours per day, 260 days each year, it would take nearly two years for me to run this same search on each partition of this data (through the ██████████ partition). This is also a conservative estimate, ██████████████████ ████████████████████████████████████ This estimate also does not account for the other demands on Facebook's personnel and resources, occasional outages, or many other variables. This estimate is also *in addition to* the two years it would take to search the ██████████████ table.

HIGHLY CONFIDENTIAL

36.   As a result, it is neither feasible nor practicable to restore all of the partitions of the ███████ and/or ██████████ tables at the same time and ██████████ ██

3.   The Burden of Production of API Call Log Data

37.   As noted, the ███████████ and ███████████ tables are stored in Facebook's internal structured data warehouse, known as "Hive." This data warehouse is not optimized to export data in a manner that can be transferred outside of Facebook. For Facebook to export and produce data from Hive tables, the data must be decompressed, broken into small pieces, and re-written to .csv files (which can be opened with Excel). This would be a manual, iterative process conducted by Facebook data scientists on the legal team. As a result, the timeline for completing an export needs to account not only for the limitations on Facebook's computing capacity, the size limitations on our mechanisms for export, and the availability of Facebook's technical resources, but also the possibility of human error, which necessitates real-time quality checks of the data downloads.

38.   Before any analysis or export can be conducted, the data must be transitioned to warm storage in order to be reviewed, analyzed, or exported. As discussed above in Paragraphs 28 and 35, I estimate it would take 5,348 hours to restore all of the partitions of ██████████ and ██████████ tables in to warm storage, which is more than 2.5 years of work, assuming I work exclusively on this task for 8 hours per day, 260 days each year, and do not run any searches to narrow the data set. Running searches on the data set would take at least an additional year.

11   **HIGHLY CONFIDENTIAL**

39.   After data has been restored into warm storage, it is necessary to decompress the data, which requires that Facebook take several additional steps, which are themselves time-consuming.

40.   First, a Facebook data scientist must write entirely new code to query the data and demarcate each partition into chunks of data that can be effectively transported to and/or received by the recipient. For instance, if the recipient intends to use Excel to analyze the data, each partition must be broken into chunks of 1 million rows each, which is the maximum number of rows that can be opened in an Excel file. When Facebook must produce Hive data in litigation, it typically produces it in Excel form. If the intended recipient—in this instance, Plaintiffs—had access to or intended to use specialized software to access the data (*e.g.*, SQL, R, Stata, SAS) and had sufficient computing power to access larger files, the limiting factor would be a combination of size of the drive used to transport the data and the maximum number of rows that could be reliably written into a .csv and reviewed for quality control. I understand from my colleagues at Facebook that the largest .csv files that we have handled and transferred reliably is 50 GB, which is roughly ███ ████████████████████████████████.

41.   Second, the code is run to export that data to a server, which would involve querying each chunk of ██████████████████ from the data warehouse and then writing the results of that query as .csv files in a decompressed flat-file format. The speed at which Facebook can export the data from the warehouse to the server varies depending on usage of its systems in the ordinary course of business, but based on testing and prior exports, Facebook's current best estimate is that it takes about 40 minutes to write a 50 GB .csv file. At this rate, it would take approximately 15 hours to extract the ██████████████ of data I identified in my

HIGHLY CONFIDENTIAL

search of the ███████ partition of the ███████ table ████████████

███████, discussed in Paragraph 27 above, to .csv format. The resulting .csv files would be

approximately █████ total, or approximately ████████████████████

████ It would then take our eDiscovery and Data Science teams about 3 minutes <u>per file</u> to

inspect the data to confirm that it was not corrupted or truncated during export. If we were to

export the ████████████ of data I identified in my search of the ████████ partition of

the ████████ table—████████████████████████████

████████████████████████, it would take approximately 1.5

hours to inspect these files. It would therefore take approximately 16.5 hours to extract and

inspect .csv files containing ████████████████████

42.   It would separately take 2 hours to extract the ████████████ data I

identified in my search of ████████████████████████████

discussed in Paragraph 34 above, to .csv format. The resulting .csv files would be approximately

████ total. It would then take our e-Discovery and Data Science teams about 3 minutes <u>per</u>

<u>file</u> to inspect the data to confirm that it was not corrupted or truncated during export. If we were

to export the ████████████ data I identified in my search of the ████████ partition

of the ████████ table—████████████████████████████

████████████████████████ it would take

approximately 12 minutes to inspect these files. It would therefore take an additional 2 to 3

hours to extract and inspect .csv files containing ████████████████

████████████

43.   Third, once the server is full, a data scientist must empty the server by transferring

the .csv files into a local encrypted hard drive or a secure transfer site. Each 50 GB .csv file

HIGHLY CONFIDENTIAL

would require 5 to 8 hours to be transferred from the Facebook servers to a physical hard drive, depending on network speeds and other variables.

44.   Decompression of the data, which is required for export of this data, results in a significant expansion in the size of the data. Within Facebook's systems, one partition of the █████████ table comprises approximately ██████████ of data and one partition of the █████████ table comprises approximately ████████ of data. It is difficult to estimate the compression ratio before the data is actually exported. Based on testing and prior exports, Facebook's current best estimate as to how big a single partition of each of these tables would be when fully decompressed to .csv format for export is around ████████ per partition for the █████████ table and █████████ per partition for the █████████ table.

45.   ██ ███ As a result, the second and third steps of this process would have to be run a minimum of 19,250 separate times for each partition of the █████████ table. Similarly, the second and third steps of this process would have to be run a minimum of 4,750 separate times for each partition of the █████████ table. It is not possible to expand the number of available servers without investing significantly in additional resources or interrupting other business and legal functions.

46.   The decompressed volume of a single partition of data from each of these tables—up to █████████████████████ table and ████████████████ for the █████████ table—is significantly larger than any hard drives or secure transfer site that Facebook has access to. Facebook's current best estimate is that exporting a single partition of the █████████ table would require it to load .csv files onto █████████ encrypted hard drives that would cost approximately $1,200 each, presuming Facebook can acquire a

14   **HIGHLY CONFIDENTIAL**

sufficient number of them. Facebook's current best estimate is that exporting a single partition

of the [REDACTED] table would require it to load .csv files onto [REDACTED] encrypted

hard drives that would cost approximately $1,200 each, again presuming Facebook can acquire a

sufficient number of them.

47.   The process must be repeated over and over until the export is complete.

Facebook estimates that it would take 8,000 days and $70,000 in hardware costs to export the

smallest, single partition of the [REDACTED] table and 2,000 days and $17,000 in hardware

costs to export a single partition of the [REDACTED] table.[1] Replicating this process for more

than [REDACTED] of each table would multiply those timing estimates proportionally.

48.   This manual, iterative process introduces the possibility of human error, which

necessitates real-time quality checks of the data downloads that take additional time and create

the potential for delay.

49.   Facebook also estimates that it will take several days for senior members of the

eDiscovery data scientist team conducting the export to prepare for and manage the coding and

export project. After the initial exported files have been created, Facebook estimates it would

take an additional month for a review team to validate the queries and data and perform

additional quality control, or longer if there are issues that require remediation.

A.   Impact on Facebook's E-Discovery Team's Resources

50.   In total, Facebook currently estimates that, even without any other responsibilities

or deadlines, it would be substantial, multi-year project to attempt to export the [REDACTED]

table and the [REDACTED] table through the process described above. The Facebook E-

Discovery team, which is responsible for assisting Facebook in-house and outside counsel in

---

[1] The shipping costs for the hard drives would add additional costs onto this project.

HIGHLY CONFIDENTIAL

active litigations and other legal matters, in addition to building and maintaining internal

infrastructure crucial to the management and preservation of data on legal hold, does not have

the time and resources required to access, analyze, and export the data in the ███████████

table and the ███████████████ table in the manner described above.

51.   Facebook cannot materially shorten this timeline by hiring new employees

because it would take time and resources to interview, select, and onboard new employees, and

any new hires would have to be sufficiently trained regarding Facebook's systems, policies, and

procedures, which is a mandatory process that takes three weeks. The API call log data export in

particular requires specialized knowledge about Facebook's systems because the system is not

built for the purpose of exporting the extremely large amounts of data at issue—knowledge that

non-Facebook personnel do not have, given the nature of the proprietary systems and databases

employed by Facebook. For the same reason, Facebook cannot simply engage third party

consultants or temporary employees to handle this data export. Nor would adding more

servers—which would require diverting them from their use in the ordinary course of business—

necessarily reduce the estimated timeline in a linear fashion, as the export still requires

downloading the data from the server, and the manual process of writing the code and

monitoring the export. All of these options—hiring new employees, hiring contractors, and

adding servers—would also be extremely costly.

4.   **Alternative Production Proposal**

52.   I understand that Facebook has proposed to produce summary data relating to API

calls to Plaintiffs in this matter instead of data from the ███████████ and ███████████

tables. ███

16   **HIGHLY CONFIDENTIAL**

███

██████████████████████████████.

53.   Specifically, I understand that Facebook has proposed to produce data from the

████████████████████ table. The ████████████████ table contains summaries of data

███

███

███

███

███

████

54.   Because the ██████████████████ table summarizes ████████████, it is

substantially less burdensome to store and analyze.

55.   The ██████████████████ table contains summary data regarding API calls

made to the Facebook Platform, ████████████████████████████████████

████████. The fields contained in this table include, among other things, ██████████

███

█████████████████████████████████   A true and correct list of the data fields

reflected in the ██████████████████ table is attached hereto as **Exhibit E**. A true and

correct sample of ████████████ from the ██████████████████ table is attached hereto as

Exhibit F.

56.   The data in the ██████████████████ table dates back to ██████████████.

57.   The ██████████████████ table is maintained in warm storage in the Hive

data warehouse. It is currently █████ in this compressed form. A single partition ██████████

HIGHLY CONFIDENTIAL

of this table is around ███ in compressed form and ███ in raw (*i.e.*, not compressed) form.

58.   It is difficult to estimate the compression ratio before the data is actually exported. Based on testing and prior exports, Facebook's current best estimate as to how big an export of data from this table dating from ████████████████ would be when fully decompressed to .csv format for export is █████

59.   Facebook is currently preparing an export of data from this table dating from ████████████████ for production to Plaintiffs in this matter and estimates it will be able to produce this data by November 30, 2021.

60.   A brief summary of the three tables discussed herein is provided below, along with relevant data points regarding the burden of production.

	████████	████████	████████
Earliest Date	████████	████████	████████
Compressed Size	████	████	████
Est. Decompressed Partition Size	████	██████	████
Est. Decompressed Size of Data Through ████	████	████	████

I declare under penalty of perjury that the foregoing is true and correct, and that I executed this Declaration on October 18, 2021, in Sausalito, California.

Mengge Ji

HIGHLY CONFIDENTIAL

Exhibit A

to Ji Declaration

Highly Confidential

REDACTED

Exhibit B

to Ji Declaration

Highly Confidential

REDACTED

REDACTED

REDACTED

REDACTED

REDACTED

REDACTED

REDACTED

Exhibit C

to Ji Declaration

Highly Confidential

REDACTED

Exhibit D

to Ji Declaration

Highly Confidential

REDACTED

REDACTED

REDACTED

REDACTED

REDACTED

REDACTED

Exhibit E

to Ji Declaration

Highly Confidential

REDACTED

Exhibit F

to Ji Declaration

Highly Confidential

REDACTED

EXHIBIT C

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

GIBSON, DUNN & CRUTCHER LLP
Orin Snyder (*pro hac vice*)
  osnyder@gibsondunn.com
200 Park Avenue
New York, NY 10166-0193
Telephone: 212.351.4000
Facsimile: 212.351.4035

Kristin A. Linsley (SBN 154148)
  klinsley@gibsondunn.com
Martie Kutscher (SBN 302650)
  mkutscherclark@gibsondunn.com
555 Mission Street, Suite 3000
San Francisco, CA 94105-0921
Telephone: 415.393.8200
Facsimile: 415.393.8306

GIBSON, DUNN & CRUTCHER LLP
Deborah Stein (SBN 224570)
  dstein@gibsondunn.com
333 South Grand Avenue
Los Angeles, CA 90071-3197
Telephone: 213.229.7000
Facsimile: 213.229.7520

Joshua S. Lipshutz (SBN 242557)
  jlipshutz@gibsondunn.com
1050 Connecticut Avenue, N.W.
Washington, DC 20036-5306
Telephone: 202.955.8500
Facsimile: 202.467.0539

Attorneys for Defendant Facebook, Inc.

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

IN RE: FACEBOOK, INC. CONSUMER PRIVACY USER PROFILE LITIGATION, This document relates to: ALL ACTIONS	CASE NO. 3:19-CV-04591-VC **FACEBOOK, INC.'S MOTION FOR A PROTECTIVE ORDER AGAINST PRODUCTION OF API CALL LOGS** Discovery Special Master Daniel Garrie, Esq.

TABLE OF CONTENTS

FACEBOOK, INC.'S MOTION FOR A PROTECTIVE ORDER AGAINST PRODUCTION OF API CALL LOGS
CASE NO. 3:18-MD-02843-VC

Gibson, Dunn &
Crutcher LLP

TABLE OF AUTHORITIES

FACEBOOK, INC.'S MOTION FOR A PROTECTIVE ORDER AGAINST PRODUCTION OF API CALL LOGS
CASE NO. 3:18-MD-02843-VC

Gibson, Dunn &
Crutcher LLP

1

I.   INTRODUCTION

2   Facebook respectfully seeks a protective order confirming that two data tables that log the

3 granular details of trillions of interactions between apps—including Facebook's own apps—and

4 Facebook's servers for more than three years are not reasonably accessible sources and that it will not

5 be required to restore, analyze, and produce the data in the two tables.

6   Millions of apps interact with Facebook every day as Facebook's more than 2 billion users play

7 games, read news, share music, and connect with each other. Facebook has explained to Plaintiffs for

8 more than one year that Facebook cannot, as a technical matter, produce two data tables detailing every

9 single *attempt* an app makes to interact with Facebook (the two tables at issue show only the types of

10 data apps requested, not what data the apps actually received). The volume of compressed data

11 contained in Facebook's logs of those interactions amounts to ███████████████████—the

12 equivalent of the storage found on approximately ██████ laptops[1] (by more conventional measures,

13 the volume of data in the two tables is equivalent to ███████ printed pages). And Facebook cannot

14 simply transfer that compressed data onto thousands of hard drives,[2] deliver those hard drives to

15 Plaintiffs, and call it a day: the tables at issue are currently retained in a proprietary format—highly

16 compressed, non-readable, and █████████████—that cannot easily be transferred out of Facebook.

17 Even Facebook, which has some of the most sophisticated data-storage capacities and data-processing

18 tools in the world, cannot restore, analyze, and produce that volume of compressed data without

19 undertaking a lengthy multi-year process.

20   Nor should Facebook have to. Logs recording the details of trillions of interactions that apps

21 have attempted to have with Facebook's servers after this case was filed venture far beyond the types

22 of data-sharing at issue in this lawsuit. On top of that, Facebook is able to produce data from a more

23 accessible table that summarizes ██████████████████████████████████. This

24 smaller table will give Plaintiffs access to the types of information they seek, despite its irrelevance to

25 their claims.

26

---

27 [1] This estimate is based on the hypothetical storage capacity of 256 GB per laptop computer.

28 [2] One petabyte is typically described as the equivalent of 500 billion pages of standard printed text.
Of course, as the Special Master well knows, the data in the tables at issue would never be printed
onto standard pages—this comparison is included only to illustrate the volume of data in the tables.

1

Gibson, Dunn &

Crutcher LLP

1   Facebook has repeatedly asked Plaintiffs why they want any of the data in these tables (which,

2   again, show only the types of data apps requested, not what data they actually received) so that

3   Facebook can attempt to identify and provide some reasonable volume of responsive information that

4   might actually be relevant to this case. The only answer Plaintiffs have provided is that they plan to

5   use the tables "as either a proxy or the actual measure for identifying how much data is being transferred

6   to third parties." Ex. 5 to Falconer Decl. But none of the tables at issue—neither the two inaccessible

7   tables nor the smaller summary table—show whether data was actually transferred to third parties. And

8   they certainly do not show whether "sensitive information" was shared with apps through friend sharing

9   or partnership arrangements without consent, which is the only type of data sharing at issue in this case.

10   Even if those types of information could, in theory, be reconstructed from the tables at issue here,

11   demanding ███████████████ of raw, granular data that Plaintiffs cannot even store, much less

12   analyze, is not a viable approach to discovery. And yet, after more than a full year of discussions,

13   Plaintiffs refuse to narrow or clarify their request to capture anything short of all records of all apps'

14   attempted interactions with Facebook's servers.

15   Plaintiffs either do not understand what the data they demand actually is, or they do and are

16   demanding massive volumes of inaccessible data solely for the improper purpose of imposing an

17   unreasonable burden on Facebook. Either way, Facebook cannot produce the two tables at issue, and

18   the Special Master should relieve it of any obligation to do so in response to Plaintiffs' RFPs and

19   Interrogatories. Facebook respectfully seeks a protective order that the data in the two tables titled

20   ██████████████ ("Mobile Table") and ██████████████ ("Web Table") is not reasonably

21   accessible because of undue burden and cost and that Facebook is protected from producing them.

II.   BACKGROUND

23   **A.   Plaintiffs refuse to confine discovery to the scope of the case and the bounds of Rule 26.**

24   Since Judge Chhabria issued his motion-to-dismiss order in September 2019, the parties have

25   been mired in an aimless fact-discovery quagmire, and their disputes have drifted further and further

26   away from the case Judge Chhabria allowed to move forward. This is a result of Plaintiffs' refusal to

27   approach discovery in a reasonable and targeted manner and increasingly desperate attempts to keep

28   this case stalled in discovery purgatory.

FACEBOOK, INC.'S MOTION FOR A PROTECTIVE ORDER AGAINST PRODUCTION OF API CALL LOGS
CASE NO. 3:18-MD-02843-VC

1   Plaintiffs have now served 90 requests for production—including 26 new requests that they

2   served just a few weeks ago. Nearly two-thirds of Plaintiffs' 90 requests indiscriminately seek "all"

3   materials about nearly every aspect of Facebook's business, without any respect for Judge Chhabria's

4   motion-to-dismiss order, narrowing the claims that may proceed, or the discovery stay he imposed on

5   the claims he did not allow to move forward. But when Facebook has tried to work with Plaintiffs to

6   narrow the scope of these requests (including the requests at issue in this motion), Plaintiffs have been

7   categorically unwilling to agree to any limitations on the scope of their requests or otherwise

8   compromise in the meet-and-confer process. As a result, the discovery process in this case has been

9   one of the most costly and burdensome Facebook has ever endured.

10   **B.   Plaintiffs demand all API call data of any kind.**

11   The dispute at issue here follows the same pattern as most other disputes in this case. Plaintiffs'

12   RFP 37 asks for "all records or logs of API Calls made by Third Parties for the Content and Information

13   of Friends of Installing Users." Ex. 1 to Falconer Decl. Plaintiffs have also issued two interrogatories

14   that seek granular detail about API calls made by "Third Parties" and "Business Partners." Ex. 3 to

15   Falconer Decl.

16   Plaintiffs' demand for "all records or logs of API Calls," refers to Application Programming

17   Interfaces, which are commonly referred to as "APIs." APIs are not unique to Facebook. They are a

18   standard industry programming tool that allow computers and computer programs to talk to or send

19   information to each other. Facebook has thousands of APIs that serve different purposes, and millions

20   of apps—including Facebook's own apps—interact with these APIs every second of every day to

21   provide services to users. Each connection to an API is referred to as a "call." *See* Ji Decl. ¶ 5.

22   Facebook employs APIs to allow Facebook's more than two billion dispersed users to be able to access

23   their Facebook profiles and those of their friends from local devices using Facebook's own apps (*e.g.*,

24   when a user pulls up their profile on a phone, the Facebook app calls Facebook's central servers through

25   APIs to access the photos, posts, and other information displayed on that profile). *See id.* ¶ 5. Facebook

26   also uses APIs to allow third parties to connect with its servers for limited purposes, including for

27   providing certain user products or experiences (*e.g.*, a website can connect to Facebook's APIs to verify

28   a user's identity rather than requiring the user to create a specific account for that site). *See id.* ¶ 6.

3

FACEBOOK, INC.'S MOTION FOR A PROTECTIVE ORDER AGAINST PRODUCTION OF API CALL LOGS
CASE NO. 3:18-MD-02843-VC

1   While Plaintiffs purport to limit their requests for API call records to records that involve an

2 app accessing user data through the user's friends, there is no way to filter for logs of API calls in this

3 way using these two data tables. *See* Ji Decl. ¶ 19. As a result, Plaintiffs now seek to impose the

4 enormous burden and cost on Facebook of restoring, analyzing, and producing massive volumes of

5 raw, granular data detailing every single one-way interaction any app, including Facebook apps, have

6 had with any API. Plaintiffs are doing so despite the fact that these tables do not show the responses

7 from Facebook to the calls and even though there is no way for Plaintiffs to use this data to ascertain

8 how much data was transferred or whether it was done within the scope of users' permissions.

9 **C.   Facebook maintains multiple responsive tables.**

10   Facebook maintains internal logs of the "API calls" made to its servers, which can serve a

11 variety of purposes, ██

12 For instance, the number of API calls increased dramatically after stay-at-home orders were issued

13 during the COVID-19 pandemic, which required Facebook to expand its data infrastructure resources

14 to meet that need. These tables log calls made to Facebook's APIs, but do not reflect whether any data

15 was returned in response to a call for data. There are a variety of reasons why data requested through

16 an API call might not be returned to an app—██

17 ██

18 ██████████████████████████████████████   *See* Ji Decl. ¶ 14.

19   Facebook's two ██████████████████ logs of API call activity are kept in two data tables:

20 the Mobile Table, ██

21 ██████████████████████ and the Web Table, ██

22 ████████████████████. Ji Decl. ¶¶ 8-9. The Mobile Table and Web Table reflect the number of calls

23 made to the Facebook APIs and additional granular information about those calls. The data fields

24 included in the Mobile Table include ██

25 ████████████████████████████ *See* Ex. A to Ji Decl. The data fields in the Web Table include ██

26 ██

27 ██████████████████████████ Ex. C to Ji Decl. Exhibits B and D to the Declaration of Mengge

28 Ji provide excerpts from the Mobile Table and Web Table respectively.

1   The Web Table and Mobile Table hold an astonishing volume of data, even a tiny fraction of

2 which would be impossible for anyone to host in a form amenable to meaningful analysis. Even in

3 highly compressed non-readable form, these tables comprise ███████████████ (a petabyte

4 is one *million* gigabytes) that Facebook does not have the infrastructure to simply pipe out of its data

5 warehouse. Even ████████ of highly-compressed data can be as much as ████████, which is

6 too large to export, analyze, or produce in any reasonable amount of time.

7   Consistent with Facebook controls related to data privacy, these tables are normally subject to

8 ██ retention periods. But, out of an abundance of caution, in light of Plaintiffs' refusal to narrow the

9 scope of their discovery requests in the meet-and-confer process, the tables have been preserved since

10 █████████████ Facebook preserves both tables in a highly compressed, non-readable format.

11 Facebook maintains the Mobile Table from █████████████, and it contains approximately ██

12 ██████ of data in highly compressed form. Facebook maintains the Web Table from ██████████

13 ███████, and it contains approximately ██████████ of data in highly compressed form. As explained

14 more below, processing these materials for analysis and external production would be prohibitively

15 expensive, time consuming, wasteful, and especially disproportionate given the limited relevance of

16 the data to Plaintiffs' claims.

17   Facebook also maintains a smaller table that summarizes the data in the ████████████

18 ██. Ji Decl. ¶ 53. Specifically, a table called █████████████████ ("Method Table") provides

19 summary data ██

20 ████████████. Ji Decl. ¶ 53. The data fields included in the Method Table are ██████████

21 ██

22 ████████████████. *See* Ex. E to Ji Decl. Exhibit F to the Declaration of Mengge Ji displays an

23 excerpt from the Method Table. Facebook maintains the Method Table from ████████████, and it

24 contains approximately ██████████ of data when restored into a form that could be transported out

25 of Facebook's systems. This is the equivalent of the storage capacity found on approximately ████

26 ████.[3] ██

27 ██

28

---

[3] This estimate is based on a hypothetical storage capacity of 256 GB per laptop computer.

5

Gibson, Dunn &
Crutcher LLP

1 Facebook anticipates it will be able to produce data from this table dated between ███████ by

2 the end of November, 2021.

3   Since Plaintiffs first served RFP 37 in May 2020, Facebook has explained to Plaintiffs that the

4 tables do not have the information they seek and that it would be impossible for Facebook to export,

5 process, and produce the volume of information Plaintiffs demand. Facebook has practically begged

6 Plaintiffs to explain what specific data they are looking for or to, at least, provide a narrowing principle

7 so that Facebook can identify some reasonable set of data to produce. But Plaintiffs have continued to

8 demand "all" and will not even agree that any productions of API call information will be restricted to

9 the same time-period the parties agreed to for custodial collections (2007 through 2019).

10   After nearly a year of trying to negotiate with Plaintiffs regarding the scope of their request for

11 API Call Log data, Facebook raised Plaintiffs' requests in mediation, seeking clarification of the scope

12 of Plaintiffs' request and the data they expected to be produced in response. Facebook hoped through

13 mediation to gain sufficient clarity regarding what Plaintiffs seek to allow it to offer a subset of API-

14 related data. Ex. 7 to Falconer Decl. Plaintiffs did not timely respond to Facebook's questions or offer

15 any narrowing of their request. On October 6, 2021, the mediators declared impasse.

16   **III.   ARGUMENT**

17   Courts have long recognized that the vast expansion of electronically stored information in

18 the digital age has magnified the risk of discovery abuse: "[t]he information explosion of recent

19 decades has greatly increased both the potential cost of wide-ranging discovery and the potential for

20 discovery to be used as an instrument for delay or oppression." Fed. R. Civ. P. 26(b) advisory

21 committee's note to 1993 amendment. This results in a danger that the discovery process "may be

22 perverted into an arena for economic power plays, that parties use . . . not so much to learn what the

23 facts are, but more to muscle one another." *In re Convergent Tech. Sec. Litig.*, 108 F.R.D. 328, 332

24 (N.D. Cal. 1985).

25   This dynamic is hard to overlook in this case. It has caused disputes to arise early and often.

26 Candidly, there is no reason for this dispute to be before the Special Master. The parties should have

27 been able to reach an agreement that Facebook does not need to do the impossible and the

28 impracticable— ████████████████ of irrelevant data that is currently in a highly

Gibson, Dunn &
Crutcher LLP

1 compressed, non-readable format in cold storage and that neither party has the technical ability to

2 host in a usable format, much less process or analyze. But Plaintiffs have refused to agree to ***any***

3 narrowing or clarification of RFP 37 or its other requests for API call log data. So Facebook has no

4 choice but to respectfully seek an order from the Special Master that (1) the Web Table and Mobile

5 Table are not reasonably accessible because of undue burden and cost under Rule 26(b)(2)(B), and

6 (2) Plaintiffs have not shown good cause for the Tables' production given their de minimis (if any)

7 relevance and the availability of alternate, summary data. *See* Fed. R. Civ. P. 26(c)(1) (discovery

8 order may issue to "protect a party or person from annoyance, embarrassment, oppression, or undue

9 burden or expense," including orders "forbidding the disclosure or discovery")

10 **A.   The API Call Logs and the data they contain are not reasonably accessible because of
       undue burden and cost.**

11   Rule 26(b)(2)(B) relieves parties from any obligation to "provide discovery of electronically

12 stored information from sources that the party identifies as not reasonably accessible because of undue

13 burden or cost." "In general, [ESI is] inaccessible [if it] 'is not readily useable and must be restored to

14 an accessible state before the data is usable.'" *U.S. ex rel. Carter v. Bridgepoint Educ., Inc.*, 305 F.R.D.

15 225, 240 (S.D. Cal. 2015) (quoting *Zubulake v. UBS Warburg LLC*, 217 F.R.D. 309, 320 (S.D.N.Y.

16 2003)).

17   Because of their size, the Web and Mobile Tables present a textbook case of inaccessibility.

18   The Mobile and Web Tables represent an astonishing volume of data that taxes even Facebook's

19 state-of-the-art storage systems to maintain in its highly compressed, non-readable form. Even if

20 Facebook could get this data out of its storage systems, it is unlikely that Plaintiffs or any entity they

21 contract with would have the data storage capacity to host it, even if Plaintiffs were willing to undertake

22 that expense. And even a tiny fraction of this data is tremendously burdensome to endeavor to produce:

23 ▮▮▮▮▮▮▮▮▮▮▮ can be as much as ▮▮▮▮▮▮ (a petabyte is one *million* gigabytes) when

24 decompressed for production, the equivalent of the storage found on approximately ▮▮▮ laptop

25 computers.[4]

26   ▮▮▮▮▮▮▮▮▮▮ Plaintiffs seek discovery of the data in more than ▮▮▮▮▮ of API call

27

28 ─────────────────
[4] This estimate is based on the hypothetical storage capacity of 256 GB per laptop computer.

FACEBOOK, INC.'S MOTION FOR A PROTECTIVE ORDER AGAINST PRODUCTION OF API CALL LOGS
CASE NO. 3:18-MD-02843-VC

log data from the Mobile and Web Tables.

Before this litigation, data from the Mobile and Web Tables was deleted in the ordinary course after ███

███ Ji Decl. ¶ 16. ███████████████████████████

██

██ *See id.* ¶ 17. Once in cold storage, the data in the Logs cannot be reviewed or analyzed. *Id.* ¶ 18. There are multiple preparatory steps required to restore the data in order to analyze it, and then more steps to export it for production. And these steps cannot be taken all at once: the limitations on Facebook's servers and computing capacity (which are industry-leading) mean that Facebook must restore and analyze these tables ████ ██████████████████████. *Id.* ¶ 19.

In order to analyze this data, Facebook data scientists must identify ████████████████████ and restore it to "warm storage"—a more accessible form of data storage that makes the data readable. Ji Decl. ¶¶ 20. It takes about one day of processing to restore 'just' ████████ of data from the Tables to warm storage—there are presently more than ████████████ in cold storage between the Mobile and Web Tables and more than ████████████, which must be analyzed in sets between ████████████ at a time. *Id.* ¶ 8, 10, 18, 21. As a result, even just making this data accessible for analysis would require a laborious process of "restoring" partitions of data to a more readily accessible form of storage, analyzing those partitions, deleting that data from warm storage, and then starting on the next partitions. *See id.* ¶¶ 37-46. This process would need to be repeated ████████████ over many years in order to search all of the data in these tables. *Id.* ¶ 47.

The segmentation and size of this data also means that Facebook cannot quickly or efficiently run searches across ████████████ data from these tables at one time, ████████████████████ ████ *Id.* ¶ 19. Rather, because the Tables are partitioned █████ each partition must be individually decompressed and searched. It could take several hours just to run a single search ████████████ of data. *See id.* ¶¶ 27-28, 34-35. These limitations essentially negate any advantage created by searching or filtering the data, as a data scientist must still work through each partition piecemeal.

A Facebook data scientist, Mengge Ji, undertook a sampling of the Web Table and the Mobile

8

Table in order to estimate how long it would take to run a simple search ████████████ ████████. *Ji Decl.* ¶¶ 22-36. She estimated it would take approximately ***four years of dedicated full-time work*** for her to restore one partition, run searches across that restored partition, and then repeat for each of ██████████████ of these tables, not including the additional time it would take to export data for production. *Id.* ¶ 35. In the real world of unexpected glitches, delays, human error, and the many other demands on Facebook's servers and processing power (most importantly, providing the Facebook family of products) this process would take much longer.

The foregoing describes only the onerous process of restoring and analyzing the extremely large quantities of data in the Tables within the proprietary Facebook system designed to house it. To produce even portions of the Tables to Plaintiffs, numerous additional obstacles would have to be overcome. Following restoration to warm storage, the data would have to be decompressed, broken into small pieces, and rewritten to a usable form, such as a .csv file that can be opened in a Microsoft Excel spreadsheet or ingested into a specialized database. *Ji Decl.* ¶¶ 40-44. It would then need to undergo quality control and be manually transferred to a mechanism that can be used to transfer the information to Plaintiffs—likely dozens of maximum capacity hard drives, costing more than a thousand dollars each, to transport ██████████████. *Id.* ¶¶ 45-47.

Even that is an oversimplification. In fact, a Facebook data scientist would have to develop entirely new code to query the data and write it into .csv files of a designated size, depending on how Plaintiffs intend to receive and analyze the information. *Ji Decl.* ¶ 40. This effort would be further complicated by any need to redact or exclude ██████████████ from the Tables. If Plaintiffs' counsel intended to review these materials in Excel, these files would need to be ████████ each (the maximum size for Excel spreadsheets). Facebook data scientist Mengge Ji found that just one partition ██████████████ of the Mobile Table contained more than ████████████ which would result in ████████ Excel files of maximum size. *Id.* ¶¶ 25-27. Assuming each row does not extend past the width of a single sheet of standard 8.5x11 copy paper, printing out the spreadsheets from this lone partition at 40 rows per page would result in a stack of paper over ████████████.

By restricting the data to a list of specific apps, Ms. Ji was able to reduce the volume to about ████████████—still nearly ████████████ Excel files. *Ji Decl.* ¶ 27. Since it takes about two

minutes to write a million rows into a .csv file, it would take almost ███████ to create this restricted data set ████████ of data from a single table. Inspecting the files for the occasional corruption or inadvertent truncation that would inevitably occur would take a data scientist another eight days to complete if this data were to be provided in Excel and Facebook's personnel worked non-stop. *See id.* ¶¶ 41-42. While this timeline may be shortened by providing Plaintiffs with larger files (too large to be opened in Excel), many other barriers to production would remain.

After being decompressed to .csv files, this enormous volume of data (already analyzed, filtered, and written into a form usable outside Facebook) must be transferred to Plaintiffs somehow. A data scientist must compress the .csv files and transfer them to a local encrypted hard drive or secure transfer site. Ji Decl. ¶ 46. Given that a single partition from either of the Tables is significantly larger than any standard hard drives or secure transfer site to which Facebook has access, dozens of massive (and expensive) hard drives would have to be purchased, loaded, and shipped, requiring additional outlays of time and money before they can finally be delivered to Plaintiffs' counsel. *Id.* ¶¶ 46-47. All of this to ██ ██████ And, in reality, the burden and expense would be much higher. Substantial resources would be dedicated to the management and quality control of this process. Outages and technical difficulties could occur. And Facebook employees would only be expected to work standard work hours, not around the clock.

By even conservative estimates, endeavoring to just search the Mobile and Web Tables would take ***more than four years*** of work. Preparing all of that data for export would add additional years to that timeline (and purchasing the hardware needed to export this data out of Facebook would be tremendously expensive).

There can be no question that the data tables are not reasonably accessible because of undue burden and cost within the meaning of Rule 26(b)(2)(B).

B.   Plaintiffs cannot show good cause for production of inaccessible API Call Log data.

Because the Web Table and Mobile Table are not reasonably accessible, Facebook cannot be required to produce these tables unless Plaintiffs "show[] good cause, considering the limitations of Rule 26(b)(2)(C)." Fed. R. Civ. P. 26(b)(2) advisory committee's note to 2006 amendment ("The

10

requesting party has the burden of showing that its need for the discovery outweighs the burdens and costs of locating, retrieving, and producing the information.") Plaintiffs cannot come close to making the requisite showing. Under Rule 26(b)(2)(C), the court must limit discovery that "is outside the scope permitted by Rule 26(b)(1)" or "can be obtained from some other source that is more convenient, less burdensome, or less expensive." Tables showing API calls for data ██████████ (after Plaintiffs filed this case) ███████ have nothing to do with Plaintiffs' allegations that Facebook shared Plaintiffs' "sensitive information" without consent through friend sharing or certain business arrangements long before this case was filed. And to the extent the tables have any nominal relevance, a separate table summarizing the data ██████████████████ is available, and this table—███████████████ ███—covers the relevant time period.

1. The data in the Mobile and Web Tables is outside the scope of discovery permitted by Rule 26(b)(1).

Rule 26 was amended in 2015 to emphasize and mandate proportionality in discovery. "No longer is it good enough to hope that the information sought might lead to the discovery of admissible evidence. . . . Instead, a party seeking discovery of relevant, non-privileged information must show, before anything else, that the discovery sought is proportional to the needs of the case." *Gilead Scis., Inc. v. Merck & Co, Inc.*, 2016 WL 146574, at *1 (N.D. Cal. Jan. 13, 2016). Discovery must be both relevant and proportional, considering, among other factors, "the importance of the discovery in resolving the issues, and whether the burden or expense of the proposed discovery outweighs its likely benefit." Fed. R. Civ. P. 26(b)(1).

Plaintiffs cannot possibly show that the Web Table or Mobile Table is important to resolving any live claim or defense in this case. In his motion-to-dismiss order, Judge Chhabria recognized that this case arose out of an incident that occurred nearly seven years ago: a single app developer's unauthorized sale of data to Cambridge Analytica in 2015. That data was obtained in 2013 through a capability called "friend sharing," which Facebook users consented to as early as 2007 and which Facebook deprecated in 2015, long before this case was filed in 2018.

Consistent with that origin story, Judge Chhabria identified "four categories" of allegations that he would allow to move forward—all of which concern sharing data through friend sharing or

FACEBOOK, INC.'S MOTION FOR A PROTECTIVE ORDER AGAINST PRODUCTION OF API CALL LOGS
CASE NO. 3:18-MD-02843-VC

1   partnerships through which Facebook outsourced "the time, labor, and money required to build

2   Facebook's Platform on different devices and operating systems." Dkt. 298 at 6-10. Judge Chhabria

3   also made clear that each category concerns only Facebook's sharing of "sensitive user information"

4   without consent, *id.* at 6, which he described as "substantive and revealing content that users intended

5   only for a limited audience [*i.e.*, their friends], such as their photographs, videos they made, videos

6   they watched, their religious and political views, their relationship information, and the actual words

7   contained in their messages." *Id.* at 1; *see also id.* at 7, 13, 17.

8   The earliest data in the Web Table and Mobile Table goes back ████████████████ That is

9   *after* Plaintiffs filed this lawsuit and years after the types of data-sharing at issue in this case—friend-

10   sharing and the vast majority of the partnerships "to build Facebook's Platform on different devices

11   and operating systems"—were discontinued. *See* Dkt. 298 at 6-8.

12   Further, the tables do not contain any information that would allow one to identify whether

13   sensitive data was shared with a third party with or without consent. Although data in the tables could

14   potentially indicate whether a particular app *requested* data, the Logs do not identify whether the app

15   actually *received* the requested data. Ji Decl. ¶ 14. For example, if the user's privacy settings

16   prohibited disclosure of the requested data, the data would not be provided to the app. It is also

17   impossible to identify from the data in these tables whether any of the data requested was associated

18   with a user that did not expressly authorize sharing that data with the application requesting it.

19   Plaintiffs indicated previously that API call data is relevant to this case "as either a proxy or the

20   actual measure for identifying how much data is being transferred to third parties." Ex. 5 to Falconer

21   Decl. Neither the Web Table nor the Mobile Table provide insight into "how much data is being

22   transferred to third parties." They show only which API calls were made—not what data (if any) was

23   returned. In any event, the volume of API calls Facebook receives **_today_** has nothing to do with this

24   lawsuit. To the extent Plaintiffs are seeking unprecedented volumes of data at enormous expense to

25   Facebook in order to rifle through Facebook's files in the hopes of finding some other case to litigate,

26   Plaintiffs should be reminded that Rule 26 confines discovery to information relevant to "the actual

27   claims or defenses surviving [dismissal]." *In re Ashworth, Inc. Sec. Litig.*, 2002 WL 33009225, at *4

28   (S.D. Cal. May 10, 2002); *Brown v. Stroud*, 2010 WL 3339524, at *2–4 (N.D. Cal. Aug. 24, 2010)

FACEBOOK, INC.'S MOTION FOR A PROTECTIVE ORDER AGAINST PRODUCTION OF API CALL LOGS
CASE NO. 3:18-MD-02843-VC

Gibson, Dunn &

Crutcher LLP

(denying request for "discovery beyond" the scope of the claims as "indicated" in the court's order regarding a motion to dismiss).

The Web Table and the Mobile Table have no bearing on any live claim or defense, but producing any portion of those data tables would be an enormous and disproportionate burden. Even merely storing the data behind multiple layers of compression is taxing Facebook's data storage system—one of the largest and most sophisticated in the world. The storage alone costs Facebook more than ███████████. And although Facebook's data scientists have developed some ideas for how it might be done in theory—over the course of years and at extraordinary cost—in reality, the challenges of actually accessing, analyzing, and producing the data are so numerous and substantial that there is a likelihood that Facebook would never be able to produce the Web Table and Mobile Table even if it dedicated millions of dollars to doing so.

Production of the Web Table and Mobile Table is not proportionate to the needs of the case. *See U.S. ex rel. McBride v. Halliburton Co.*, 272 F.R.D. 235, 241 (D.D.C. 2011) (the discovery sought "cannot possibly be justified when one balances its cost against its utility.")

2.   Facebook is willing to produce alternative sources of information about the contents of the API Call Logs.

Burdensome production of data from the Mobile and Web Tables is also unwarranted here, because another table is available—specifically, the Method Table—that contains summaries of the data in the ███████████████. Facebook cannot conceive of any information Plaintiffs could possibly need from the Mobile or Web Tables that could not be obtained from this less burdensome source. Unlike the Mobile and Web Tables, the Method Table also has data from the relevant time period, as the data in the Method Table dates back to ███████████.

Facebook is preparing for production and intends to produce by November 30, 2021 the Method Table from ██ ██████████████████ Although it is difficult to predict how much the data set will expand when decompressed for production, Facebook estimates that the ██████████ from the Method Table

will comprise ██████ when prepared for production.[5]

The Method Table represents an enormous amount of data regarding API calls made to Facebook's servers over the course of nearly a decade. In light of this offer of production, Plaintiffs have absolutely no need to insist upon the production of the inaccessible and irrelevant data sets in the Web and Mobile Tables.

IV.   CONCLUSION

Facebook respectfully requests that the Special Master grant Facebook's Motion for a protective order.

---

[5] Facebook reserves its right to seek cost-shifting if and as appropriate under Fed. R. Civ. P. 26(b), particularly if Plaintiffs seek additional call log data beyond that which is available in the Method Table.

14

Gibson, Dunn & Crutcher LLP

1   Dated: October 18, 2021   **GIBSON, DUNN & CRUTCHER, LLP**

2   By: */s/ Russ Falconer*

3   Orin Snyder (*pro hac vice*)
osnyder@gibsondunn.com
200 Park Avenue

4   New York, NY 10166-0193
Telephone: 212.351.4000

5   Facsimile: 212.351.4035

6   Deborah Stein (SBN 224570)
dstein@gibsondunn.com

7   333 South Grand Avenue
Los Angeles, CA 90071-3197

8   Telephone: 213.229.7000
Facsimile: 213.229.7520

9

10   Joshua S. Lipshutz (SBN 242557)
jlipshutz@gibsondunn.com
1050 Connecticut Avenue, N.W.

11   Washington, DC 20036-5306
Telephone: 202.955.8500

12   Facsimile: 202.467.0539

13   Kristin A. Linsley (SBN 154148)
klinsley@gibsondunn.com

14   Martie Kutscher (SBN 302650)
mkutscherclark@gibsondunn.com

15   555 Mission Street, Suite 3000
San Francisco, CA 94105-0921

16   Telephone: 415.393.8200
Facsimile: 415.393.8306

17

18   Russ Falconer (*pro hac vice*)
rfalconer@gibsondunn.com

19   2001 Ross Avenue, Suite 2100
Dallas, TX 75201-2911

20   Telephone: 214.698.3100
Facsimile: 214.571.2900

21   *Attorneys for Defendant Facebook, Inc.*

22

23

24

25

26

27

28

Gibson, Dunn &
Crutcher LLP

EXHIBIT D

Derek W. Loeser (admitted *pro hac vice*)
KELLER ROHRBACK L.L.P.
1201 Third Avenue, Suite 3200
Seattle, WA 98101
Tel.: (206) 623-1900
Fax: (206) 623-3384
dloeser@kellerrohrback.com

Lesley Weaver (Cal. Bar No.191305)
BLEICHMAR FONTI & AULD LLP
555 12th Street, Suite 1600
Oakland, CA 94607
Tel.: (415) 445-4003
Fax: (415) 445-4020
lweaver@bfalaw.com

Plaintiffs' Co-Lead Counsel

Additional counsel listed on signature page

**UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION**

IN RE: FACEBOOK, INC. CONSUMER PRIVACY USER PROFILE LITIGATION	MDL No. 2843 Case No. 18-md-02843-VC-JSC
This document relates to: ALL ACTIONS	**PLAINTIFFS' OPPOSITION TO FACEBOOK'S MOTION FOR A PROTECTIVE ORDER TO STOP PRESERVING CERTAIN API CALL LOGS** Judge: Hon. Vince Chhabria Hon. Jacqueline Scott Corley Special Master Daniel Garrie Courtroom: 4, 17th Floor JAMS Ref. No.: 1200058674 ORAL ARGUMENT REQUESTED

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

TABLE OF CONTENTS

1

TABLE OF AUTHORITIES

2

Cases

8

Rules

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

OPP. TO FACEBOOK'S MOT. FOR   2   MDL No. 2843
PROT. ORDER RE API LOGS   CASE No. 18-MD-02843-VC-JSC
                                              JAMS REF. NO.: 1200058674

1   **I.   INTRODUCTION**

2   The Special Master should deny Facebook's motion for a protective order permitting it to

3   destroy the Mobile and Web Tables, which contain relevant information about API calls by third

4   parties ██

5   ███████

6   Plaintiffs appreciate the information Facebook provides regarding these tables and its new

7   disclosure of the existence of a Method Table. Facebook's motion and, in particular, the supporting

8   Ji Declaration provide exponentially more information than Facebook has provided about API call

9   logs during the past three years of litigation.

10   But Facebook's brief and supporting materials do not establish good cause for permitting it

11   to destroy potentially probative evidence. Nor do they provide the information necessary to

12   ascertain the benefit of preservation. Facebook's brief is based on a strawman—that Plaintiffs seek

13   the production of the entirety of the Mobile and Web Tables. They don't. Plaintiffs have always

14   been willing to target the production of data from Facebook's API call logs and have repeatedly

15   asked Facebook to disclose information about them so that they could understand how relevant

16   information could be extracted, sampled, or snapshotted. Plaintiffs do not want Ji to spend four

17   years extracting and producing ██████████ data.

18   More pertinent, Facebook's submission raises a series of questions that must be addressed

19   before the Special Master has sufficient information to determine whether Facebook should be

20   permitted to destroy the information in these tables. Ji provides considerable information about how

21   the information in the Mobile and Web Tables is currently preserved, but little specificity about

22   what the information comprises. Similarly, Ji provides a high-level summary of the information in

23   the Method Table, but too few details to understand the contents. After waiting three years for

24   Facebook to disclose any information about API call logs, it's troubling that Facebook is providing

25   meaningful but limited information about them for the first time in support of a motion seeking

26   leave for their destruction.

27

28

OPP. TO FACEBOOK'S MOT. FOR   3   MDL No. 2843
PROT. ORDER RE API LOGS   CASE No. 18-MD-02843-VC-JSC
                                                                                   JAMS REF. No.: 1200058674

1   In sum, Facebook's motion is at best premature, and at worst seeks to handicap Plaintiffs'

2   ability to prove their claims by seeking leave to delete what may be the only remaining sources of

3   certain relevant data. ██

4   ██

5   ████████████████   Plaintiffs propose that the Special Master require Facebook to answer a series

6   of specific questions listed below. Plaintiffs further propose that, once Facebook has provided

7   written answers to the questions, the Special Master convene a hearing during which he will meet

8   *ex parte* with each side and its experts, going back and forth between the two "rooms" for the

9   purpose of facilitating the provision of further information about the Mobile and Web Tables, the

10   Method Table, and other sources of API call information.

11   **II.   RELEVANT BACKGROUND**

12   Unfortunately, it is necessary to respond to Facebook's misrepresentation of relevant

13   history regarding Plaintiffs' requests for production related to API call logs. The headline is that,

14   while Plaintiffs believe Facebook's provision of information about the Mobile, Web, and Method

15   Tables is a substantial and meaningful step forward, all of the information Facebook now provides

16   for the first time should have been provided years ago. Indeed, the information Facebook now

17   provides includes information it previously told Plaintiffs did not exist.

18   ***Rule 26(f) Conference***. On November 11, 2019, the parties held their Rule 26(f)

19   conference. Among other things, Rule 26(f) requires the parties to "discuss any issues about

20   preserving discoverable information." ██

21   ██

22   ██ Ex. 1 at 1.[1] Though

23   Facebook now moves for a protective order that would permit it to permanently delete the Mobile

24   and Web Tables, it did not identify any issues associated with their preservation during the Rule

25   26(f) conference.

26

27   _____

28   [1] Except where specified otherwise, all "Ex." references are to the Melamed Declaration filed
     herewith.

1   ***RFP No. 37***. On May 6, 2020, Plaintiffs served Request for Production No. 37, which seeks

2 "all records or logs of API Calls made by Third Parties for the Content and Information of Friends

3 of Installing Users." Falconer Decl., Ex. 1 at 10.

4   On June 19, 2020, Facebook issued responses and objections related to RFP No. 37.

5 Facebook objected that it "does not have documents identifying . . . whether API calls were made

6 that sought data relating to the friends of a particular app's users." *Id*., Ex. 2 at 7. Facebook indicated

7 that it was "willing to meet and confer with Plaintiffs regarding the documents being sought by this

8 Request, their relevance to the Plaintiffs' claims (if any), and what documents Facebook could

9 reasonably produce proportionate to the needs of the case." *Id*. at 8. On July 2 and 20, 2020 meet

10 and confers, Facebook asserted that API call logs do not contain information reflecting third-party

11 access of users' information and that the logs are not "meant for human consumption." Ex. 2 at 1-

12 2. However, Facebook's motion and supporting papers now evidence that Facebook does appear to

13 have documents reflecting such information. *See* Opp. at 4 (describing the data fields included in

14 the Mobile and Web Tables).

15   ***Interrogatory No. 13***. On July 16, 2020, Plaintiffs issued Interrogatory No. 13, which

16 requested that Facebook identify, for each API that third parties could use to access restricted

17 content and information of users' friends, specific information including data fields the API could

18 access, the number of calls it received each month, the volume of data it returned each month, the

19 number of friends whose information was accessed, the identity of each third party that was allowed

20 to use it, and the time period of such use. Falconer Decl., Ex. 3 at 8.

21   On November 20, 2020, Facebook issued amended objections and responses to this

22 interrogatory. It stated it "continue[d] to investigate what information it can produce in response

23 to" Plaintiffs' request for "over 10 years of monthly data regarding the number of calls certain APIs

24 received, the volume of data returned, and the numbers of users whose data was accessed." Falconer

25 Decl., Ex. 4 at 59. It did not update that investigation until August 2021, when it first disclosed the

26 existence of (and almost nothing else about) the Mobile and Web Tables in the context of the

27 parties' ongoing discovery mediation process.

28

1   ***Interrogatory No. 15***. On July 16, 2020, Plaintiffs issued Interrogatory No. 15, which seeks

2   certain information about APIs used by Business Partners that had the ability to access restricted

3   content and information of users' friends. Among other things, Plaintiffs requested that Facebook

4   identify the number of calls each Business Partner made to and the volume of data transferred from

5   each such API on a monthly basis, and the number of users' friends whose information had been

6   accessed via these APIs.

7   On November 20, 2020, Facebook issued amended objections and responses to this

8   interrogatory. Facebook identified certain APIs for a limited subset of Business Partners. (As the

9   Special Master is aware, the parties are at impasse and have submitted briefing concerning the

10   disputed definition of that term.) But Facebook objected that answering the interrogatory's request

11   for the number of monthly calls and volume of data by API and Business Partner would require "a

12   compilation of information that is not reasonably available at this stage in the case and that would

13   be unduly burdensome for Facebook to provide." Falconer Decl., Ex. 4 at 70. Information disclosed

14   for the first time in support of Facebook's instant motion suggests that the Method Table, which it

15   now promises to produce, is likely to provide at least some information responsive to Plaintiffs'

16   request, and that the Mobile and Web Tables also contain such information.

17   The parties met and conferred about these discovery requests on July 2, 2020, December

18   15, 2020, and January 5, 2021. Each time, Facebook has insisted that Plaintiffs narrow their

19   requests. Each time, Plaintiffs requested information about the API call logs that would enable them

20   to do so. Each time, Facebook refused to provide any substantive information about its API call

21   logs but instead reiterated its generic objection on the basis of burden.

22   Indeed, Plaintiffs have served discovery specifically aimed at information they need to have

23   a meaningful discussion about how Facebook maintains API call logs and how that information is

24   made accessible. For instance, Plaintiffs' RFP No. 39 seeks Facebook's internal manuals,

25   schematics, and other documents describing the content, structure, or organization of the databases

26   containing information about Facebook Users. Ex. 3 at 2.

27

28

1   On April 6, 2020, Plaintiffs reiterated their request for these materials. "Plaintiffs seek

2   manuals or other documents describing Facebook's architecture that are commonly used to train

3   developers or give them an understanding of the schema of Facebook's systems." Ex. 4 at 2. In July

4   20, 2020 meet and confer, Plaintiffs again sought documents responsive to Request for Production

5   No. 39 that would have assisted them in meaningfully narrowing and providing more specificity in

6   their request for production of API call logs. Ex. 2 at 2; *see* Ex. 5 at PwC_CPUP_FB00021549,

7   550 (onboarding agenda for developers identifying materials Facebook has not produced related to

8   "Data: Logging, Monitoring and Analysis" and other relevant topics). Despite these entreaties,

9   Facebook did not provide any internal manual, schematic, or similar documents showing how it

10   maintains data in API call logs.

11   Plaintiffs continued to seek information about Facebook's API call logs during discovery

12   mediation. On July 20, 2021, Facebook asked Plaintiffs to "clarify the basis of their requests for

13   API call logs, so that Facebook is able to consider whether there is a reasonable set of alternate data

14   it can realistically produce that will satisfy Plaintiffs' needs and be proportionate to the needs of

15   the case." Falconer Decl., Ex. 5 at 4. On July 27, 2021, Plaintiffs responded by again asking for

16   information that would help them understand the information contained in the API call logs,

17   including the schemas, data dictionaries, and documents sufficient to identify the categories of

18   information collected; the methods used; a description of how the information is or can be observed,

19   collected, used, stored, and searched; and the categories of information contained within the

20   presently available logs that were not preserved or are not accessible for the full class period. *Id.* at

21   2-3.

22   On July 29, 2021, Facebook responded by noting that Plaintiffs mistakenly described

23   "terabytes," rather than "petabytes." *Id.* at 1. It further suggested the parties engage in a joint session

24   with the Special Master to discuss the issue. *Id.* On August 2, 2021, Plaintiffs acknowledged the

25   mistaken description of the volume of information at issue, reiterated the July 27 request for

26   information, and agreed that participating in a session conducted by the Special Master would be a

27   good next step. *Id.* Facebook did not respond.

28

1   On August 12, 2021, Facebook informed the discovery mediators that the parties were likely

2   at impasse regarding whether it would be required to "preserve or produce inaccessible API call

3   log data." Falconer Decl., Ex. 6 at 3. On August 18, 2021, Facebook provided additional detail

4   about the API call logs it maintained. *Id.* at 1. It stated for the first time that it maintained two API

5   call logs, the ███████████ (Mobile) and ███████████ (Web) tables. It stated that these tables

6   log individual calls to the Graph API and ██████████████████████████████████████

7   ██████████████████████ It described some of the fields in each of the tables, but not all. It did

8   not describe the values of highly relevant fields, ████████████████████████████████████

9   ██████████████████████████████████████

10   Facebook did not provide any additional information, and instead asked the Discovery

11   Mediators to declare impasse. On October 6, 2021, the discovery mediators declared impasse on

12   "API Call Logs." In sum, Facebook only disclosed the existence of any API call logs in July, only

13   provided a basic sketch of what they contained in August, and only disclosed the existence of the

14   Method Table in its opening brief ten days ago. While important, these disclosures are only a first

15   step. They lead to questions that must be answered before Facebook is permitted to destroy the only

16   sources of what seems to be highly probative evidence.

17   **III.   LEGAL STANDARD**

18   Under Federal Rule of Civil Procedure 26(c), the party seeking a protective order for

19   discovery materials must demonstrate that "good cause" exists for the protection of evidence.

20   *Phillips ex rel. Ests. of Byrd v. Gen. Motors Corp.*, 307 F.3d 1206, 1210–11 (9th Cir. 2002). Good

21   cause requires showing that "specific prejudice or harm will result if no protective order is granted."

22   *Id.* If the party makes this showing, "the court must then determine if an order is necessary by

23   balanc[ing] the public and private interests at stake." *Ground Zero Ctr. For Non-Violent Action v.*

24   *U.S. Dep't of Navy*, 860 F.3d 1244, 1260 (9th Cir. 2017) (quotations and citations omitted;

25   alterations in original).

26   Courts deny motions for a protective order where the information sought to be preserved is

27   "directly relevant to the claims and defenses in [the] case", and "proportional to the needs of the

28

OPP. TO FACEBOOK'S MOT. FOR   8   MDL No. 2843
PROT. ORDER RE API LOGS   CASE No. 18-MD-02843-VC-JSC
                                                         JAMS REF. No.: 1200058674

case." *Corel Software, LLC v. Microsoft Corp.*, No. 2:15-cv-00528-JNP-PMW, 2018 WL 4855268, at *1-2 (D. Utah Oct. 5, 2018) (denying defendant's motion for a protective order barring further retention and production of user data). Further, courts deny motions where defendant fails to provide the information necessary to "ascertain the 'benefit' of preservation." *Pippins v. KPMG LLP*, 279 F.R.D. 245, 256 (S.D.N.Y. 2012).

IV.   ARGUMENT

A.   Facebook Fails to Demonstrate That Good Cause Exists for Entry of the Protective Order

Facebook's submission sets forth a strawman burden, contending that it would take approximately 10,000 days and $87,000 in hardware costs to export single partitions of both the Mobile and Web Tables. Ji Decl. ¶ 47. There are two problems with that analysis.

First, the question on Facebook's motion is *not* the burden associated with production, but rather the burden of continued preservation. Facebook must demonstrate the specific harm or prejudice that will result without a protective order. *Ground Zero*, 860 F.3d at 1260. Thus, the relevant consideration is the harm that Facebook is suffering from the need to preserve the Mobile and Web Tables.

The only mention in the Ji Declaration of the harm Facebook is suffering is the current size of the tables— █████████ for the Mobile Table, █████████ for the Web Table—and the approximate cost of █████████ to store them. To be sure, those numbers are large. But they must be evaluated in context of the party claiming the burden. *See Corel*, 2018 WL 4855268 at *2 (considering preservation costs to Microsoft in context of its available resources denying protective order barring further retention of data). Here, Facebook built its first dedicated data storage center, with capacity to hold an exabyte (1,000 petabytes) of data in cold storage, in 2012. Since then, it has opened 17 more data centers. It has sufficient space to continue preserving the Mobile and Web Tables for now. Similarly, Facebook reported revenues of $84.2 billion in 2020. The █████████ █████████ to continue storing the Mobile and Web Tables reflects approximately █████████ revenues. The burden of continuing to preserve these tables while Facebook provides additional

1   information sufficient for Plaintiffs to evaluate whether they do contain relevant information is

2   minimal.

3   **Second**, Facebook's calculation of the time it would take to restore and produce the Mobile

4   and Web Tables reflects production that Plaintiffs do not seek. Plaintiffs have repeatedly told

5   Facebook that they do *not* want ███████████ and do *not* seek production of the entire Mobile

6   and Web Tables. It is likely that the vast majority of information contained within the Mobile and

7   Web Tables are not relevant, and Plaintiffs ask questions below in § IV.C aimed at understanding

8   how the relevant exports can be whittled down.

9   Indeed, the Ji Declaration states that restoring a ███████████ from the Mobile

10   Table—one way of limiting the burden of production—would take approximately 2 hours to restore

11   to warm storage, and another hour-plus to run a search across the partition to identify data

12   associated with non-proprietary Facebook apps. Ji Declaration ¶¶ 22-27. And while Ji declares that

13   "Facebook cannot quickly or efficiently run searches across ███████████" in the tables at

14   once, they continue by stating that "Facebook's system is optimized for running searches within an

15   individual partition." *Id.* ¶ 19. It is therefore likely that once Facebook provides more information

16   about the Mobile and Web Tables, Plaintiffs could provide further guidance regarding how to

17   generate more limited, meaningful samples from single partitions of the Mobile and Web Tables.

18   The burden of generating such samples would almost certainly be proportionate to the needs of the

19   case.

20   **B.** **Facebook Fails to Provide the Information Necessary to Ascertain the Benefits of Preservation**

21

22   **1.** **Information Related to API Calls Is Highly Probative**

23   As Facebook notes, Plaintiffs have identified API calls as relevant "'as either a proxy or the

24   actual measure for identifying how much data is being transferred to third parties.'" Mot. at 12

25   (citation omitted). Facebook responds by asserting that API call logs in general, and the Web and

26   Mobile Tables specifically, provide no insight into how much user information is being transferred

27   to third parties. *Id.* The limited information disclosed by Ji indicates that may not be accurate; it

28

OPP. TO FACEBOOK'S MOT. FOR   10   MDL No. 2843
PROT. ORDER RE API LOGS   CASE No. 18-MD-02843-VC-JSC
                   JAMS REF. No.: 1200058674

1   may be possible to use ███████████████████████████

2   ██████████████████████████████.

3   Regardless, Facebook doesn't and can't refute that API calls function as a proxy for

4   disclosure. Internal documents demonstrate Facebook's use of API calls to understand where and

5   to whom it transferred users' information. To provide just a few examples:

6   • ██

██

██

██

██

███████████████████████████████

12   • ██

██

███████████████████████████████████

15   • ██

██

██

█████████████████████████████████████

19   • █████████████████████████████████████

████████████████████████████████████

21   o ██

██

██████████████████████████

24   o ██

██

██

██

██████████████████████████

OPP. TO FACEBOOK'S MOT. FOR   11   MDL No. 2843
PROT. ORDER RE API LOGS   CASE No. 18-MD-02843-VC-JSC
                                                    JAMS REF. No.: 1200058674

- Facebook publicly noted that, in response to "the episode involving Cambridge Analytica," it had "removed a number of APIs, the channels that developers use to access various types of data." Ime Archibong, An Update on Our App Developer Investigation (Sept. 20, 2019).[2]

These examples demonstrate the relevance of API calls to Plaintiffs' case. ████████

██

████████████████ Facebook should not be permitted to permanently destroy them absent a deep and substantive discussion of their contents and their place within Facebook's data ecosystem.

2.   Plaintiffs Should Not Be Punished for Facebook's Delay

Having sandbagged for years, Facebook now identifies the purportedly urgent need to destroy what it says are the sole remaining API call logs. Facebook did not identify the Mobile or Web Tables until August 2021 and only disclosed the existence of the Method Table last week. It is astonishing that these disclosures were not made earlier. It is equally astonishing that, having so recently disclosed the Mobile and Web Tables' existence and provided such limited information about them, Facebook now seeks permission to destroy them forever. The burden associated with maintaining the logs in cold storage is more than offset by the need to understand their contents with specificity before those contents are made permanently inaccessible. As the *Pippins* court explained, "I cannot possibly balance the costs and benefits of preservations when I'm missing one side of the scale (the benefits)." 279 F.R.D. at 256.

Facebook asserts that the Mobile and Web Tables are irrelevant because they only contain ████████. Such information, however, is relevant. The relevant time period for Plaintiffs' claims includes dates for which the Mobile and Web Tables have data. The JPML consolidated and transferred the first set of cases against Facebook to Judge Chhabria on June 7, 2018. Information in the Mobile Table, which includes data from ████████, and the Web Table, which includes data from ████████, predate the initial consolidation order. Ji Decl., ¶¶ 8, 10. Moreover, Plaintiffs' claims cover the time period "from 2007 to the present." Dkt. No. 298 at 4. The complaint

---

[2] ██

1 Judge Chhabria considered was filed on February 22, 2019. He could have indicated that was the

2 end of the proposed class period, but his September 9, 2019 order defined the end of the relevant

3 time period as the present, recognizing that the end of the class period is currently defined by

4 Facebook's behavior, not a calendar.

5   In addition, Facebook's assertion that friend sharing ceased years before the time period

6 covered by data in the Mobile and Web Tables (Mot. at 12) is untrue. The FTC alleged that

7 Facebook maintained "private arrangements with dozens of developers, referred to as 'Whitelisted

8 Developers,' that allowed those developers to continue to collect the data of Affected Friends, with

9 some of those arrangements lasting until June 2018." *United States v. Facebook, Inc.*, Complaint

10 for Civil Penalties, Injunction, and Other Relief ¶ 8; *see also id.* ¶ 172; *see also* Ex. 13 at FB-CA-

11 MDL-01822488 (1/18/18 email: "[a] good number of OEM apps are still whitelisted for a

12 capability" that "giv[es] away access to their friends' email addresses"). The Mobile and Web

13 Tables will contain evidence of which apps could still call APIs for friends' data long after

14 Facebook stated those permissions had been universally revoked.

15   Finally, Facebook has still not identified the universe of API call log and call log-derived

16 data it maintains. ████████████████████████████████████

17 ██

18 ██ Facebook should disclose

19 all sources of API call log or log-related information before being permitted to destroy the Mobile

20 and Web Tables. Regardless, if the Mobile and Web Tables contain ██████████ API call

21 logs, it may be necessary to extract limited data from them to determine the extent to which the

22 available data is representative of granular call logs that no longer exist.

23   **C.   Plaintiffs' Proposal for Moving Forward**

24   Plaintiffs propose the following three-step process for Facebook to provide additional

25 information about Mobile, Web, and Method Tables, as well as any other sources of API call logs.

26   ***First***, Facebook should provide written answers to, or produce documents sufficient to

27 answer, the following questions.

28

Concerning the Mobile and Web Tables:

1. Is it possible to use the ███████████ information to estimate which API queries produced ██

2. Can the ████████ information be used to ███

Concerning the Mobile, Web, and Method Tables (these questions are aimed not only at understanding the Tables' contents, but also at whether there are ways to lessen the burden of production):

3. Can Facebook provide documentation enumerating all of the ██

4. Can Facebook provide documentation enumerating all of the possible values for ██████████ reflected in the Tables and how they should be interpreted ████████████████████████████████

5. Can Facebook provide documentation explaining each █████ identified in the three Tables and the time period during which that ████████ existed? Among other things: What were the ██

Concerning other data sources:

6. Does Facebook have a source that stores information about which users have authorized which ████████████████████████ Relatedly, does Facebook have a table that stores which users have installed which second- and third-party apps and, if so, can Facebook provide information from this table about the specific permissions that each user granted to each app (including but not limited to the dates on which each user installed and uninstalled, if they did, each app)?

7. For users that authorized second- and third-party apps to access their accounts, can Facebook provide the size of their set of friends (now and at past points in time)?

8. For all Facebook users who either installed a second- or third-party app, or who had a friend who installed such an app, can Facebook provide information about the privacy settings that would implicate the ability of all such aps to access the users'

data? For example, was each user opted in or out of data sharing with friends' second and third-party apps? Can Facebook provide the dates on which each setting was modified by each user?

Second, Plaintiffs propose that the Special Master convene a special hearing to discuss the implications of Facebook's answers to the foregoing questions, address follow-up or additional questions, and move towards a shared understanding of the logs Facebook seeks permission to destroy and the information that remains. Plaintiffs propose that this special hearing be conducted through *ex parte* conversations so that the parties are not required to disclose their experts before the Court-established July 15, 2022 deadline for doing so. Dkt. No. 706.

Third, the parties shall subsequently confer to see if they agree regarding the continued preservation or permitted destruction of the Mobile and Web Tables, or portions thereof. If the parties do not agree, the issue is returned to the Special Master's purview. He may deem the issue submitted, ask for supplemental briefing, or issue any other orders he determines are appropriate.

V.   CONCLUSION

The Special Master should deny Facebook's motion for permission to destroy the Mobile and Web Tables, which it describes as the ███████████████ logs of API call activity. Facebook has not provided sufficient information to convey the scope of information they contain, what other API call logs exist, or the true burden of exporting a relevant sample of the data they contain.

Dated: October 28, 2021

Respectfully submitted,

KELLER ROHRBACK L.L.P.   BLEICHMAR FONTI & AULD LLP

By: */s/ Derek W. Loeser*   By: */s/ Lesley E. Weaver*
Derek W. Loeser   Lesley E. Weaver

Derek W. Loeser (admitted *pro hac vice*)   Lesley E. Weaver (SBN 191305)
Cari Campen Laufenberg (admitted *pro hac vice*)   Anne K. Davis (SBN 267909)
David Ko (admitted *pro hac vice*)   Matthew S. Melamed (SBN 260272)
Adele A. Daniel (admitted *pro hac vice)*   Angelica M. Ornelas (SBN 285929)
Benjamin Gould (SBN 250630)   Joshua D. Samra (SBN 313050)
1201 Third Avenue, Suite 3200   555 12th Street, Suite 1600
Seattle, WA 98101   Oakland, CA 94607
Tel.: (206) 623-1900   Tel.: (415) 445-4003
Fax: (206) 623-3384   Fax: (415) 445-4020
dloeser@kellerrohrback.com   lweaver@bfalaw.com
claufenberg@kellerrohrback.com   adavis@bfalaw.com

1

dko@kellerrohrback.com
adaniel@kellerrohrback.com
bgould@kellerrohrback.com

mmelamed@bfalaw.com
aornelas@bfalaw.com
jsamra@bfalaw.com

2

3

Christopher Springer (SBN 291180)
801 Garden Street, Suite 301
Santa Barbara, CA 93101
Tel.: (805) 456-1496
Fax: (805) 456-1497
cspringer@kellerrohrback.com

4

5

6

7

Plaintiffs' Co-Lead Counsel

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

EXHIBIT E

1

2

GIBSON, DUNN & CRUTCHER LLP
Orin Snyder (*pro hac vice*)
   osnyder@gibsondunn.com
200 Park Avenue
New York, NY 10166-0193
Telephone: 212.351.4000
Facsimile: 212.351.4035

GIBSON, DUNN & CRUTCHER LLP
Deborah Stein (SBN 224570)
   dstein@gibsondunn.com
333 South Grand Avenue
Los Angeles, CA 90071-3197
Telephone: 213.229.7000
Facsimile: 213.229.7520

Kristin A. Linsley (SBN 154148)
   klinsley@gibsondunn.com
Martie Kutscher (SBN 302650)
   mkutscherclark@gibsondunn.com
555 Mission Street, Suite 3000
San Francisco, CA 94105-0921
Telephone: 415.393.8200
Facsimile: 415.393.8306

Joshua S. Lipshutz (SBN 242557)
   jlipshutz@gibsondunn.com
1050 Connecticut Avenue, N.W.
Washington, DC 20036-5306
Telephone: 202.955.8500
Facsimile: 202.467.0539

Attorneys for Defendant Facebook, Inc.

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

IN RE: FACEBOOK, INC. CONSUMER
PRIVACY USER PROFILE LITIGATION,

This document relates to:

ALL ACTIONS

CASE NO. 3:19-CV-04591-VC

FACEBOOK, INC.'S REPLY IN
SUPPORT OF MOTION FOR A
PROTECTIVE ORDER AGAINST
PRODUCTION OF API CALL LOGS

Discovery Special Master Daniel Garrie, Esq.

Gibson, Dunn &
Crutcher LLP

FACEBOOK, INC.'S REPLY IN SUPPORT OF MOTION FOR A PROTECTIVE ORDER AGAINST PRODUCTION
OF API CALL LOGS
CASE NO. 3:18-MD-02843-VC

TABLE OF CONTENTS

Gibson, Dunn &

Crutcher LLP

TABLE OF AUTHORITIES

Page(s)

ii

FACEBOOK, INC.'S REPLY IN SUPPORT OF MOTION FOR A PROTECTIVE ORDER AGAINST PRODUCTION
OF API CALL LOGS
CASE NO. 3:18-MD-02843-VC

Gibson, Dunn &
Crutcher LLP

FACEBOOK, INC.'S REPLY IN SUPPORT OF MOTION FOR A PROTECTIVE ORDER AGAINST PRODUCTION
OF API CALL LOGS
CASE NO. 3:18-MD-02843-VC

Gibson, Dunn &
Crutcher LLP

The motion that Facebook filed—which seeks a protective order against the *production* of data from the ███████████████ ("Mobile Table") and ███████████████ ("Web Table")—is not the motion that Plaintiffs oppose. Instead, Plaintiffs' Opposition proceeds from the false premise that Facebook seeks permission to *delete* data from the Mobile and Web Table, which it does not. In doing so, Plaintiffs ignore Facebook's actual request and fail to address Facebook's core arguments.

Most notably, Plaintiffs concede that the data in the Tables is inaccessible under Federal Rule 26(b)(2)(B) and cannot be produced, in whole or in part, without enormous burden and expense. And they do not argue that good cause exists under Rule 26(b)(2)(B) to nonetheless order the inaccessible data's production. The Special Master can and should grant Facebook's motion on this basis alone.

In an effort to avoid a ruling on Facebook's motion and in search of an opportunity to scrounge further through Facebook's records, Plaintiffs adopt a now-familiar tactic: seeking to delay resolution of this issue by claiming they do not have enough information to assess it. But the questions Plaintiffs pose do not have any bearing on the inaccessibility of the data in the Tables and would not refute the data's marginal relevance to their claims. Plaintiffs also ignore that the data Facebook will produce from the ███████████████ ("Method Table") will give them what they say they need. The Special Master should refuse to engage with these delay tactics and grant Facebook's motion.

I.   ARGUMENT

A.   Plaintiffs ignore their burden to establish good cause for production of inaccessible data.

Plaintiffs do not—because they cannot—dispute the central premise of Facebook's Motion: the API Call Log data in the Mobile and Web Tables is inaccessible within the meaning of Rule 26(b)(2)(B). Mot. at 7-10. Plaintiffs' Opposition does not address, let alone contest, the inaccessibility of the data. Plaintiffs thus concede the issue. *Pecover v. Elec. Arts Inc.*, 633 F. Supp. 2d 976, 984 (N.D. Cal. 2009) ("[P]laintiffs have effectively conceded, by failing to address the issue in their opposition"); *Brown v. DIRECTV, LLC*, 2019 WL 6604879, at *10 (C.D. Cal. Aug. 5, 2019) ("Typically, a non-movant's failure to raise an issue in an opposition . . . constitutes a waiver thereof.").

Plaintiffs further neglect their burden to show there is good cause to require production from these inaccessible Tables. Once a party establishes that ESI sought in discovery is not reasonably

accessible, the burden shifts to the party seeking the discovery to show good cause for production "considering the limitations of Rule 26(b)(2)(C)," which include whether the discovery sought "is outside the scope permitted by Rule 26(b)(1)" and/or "can be obtained from some other source that is more convenient, less burdensome, or less expensive." Mot. at 11 (citing Fed. R. Civ. P. 26(b)(2)(B)).

Rather than address this framework, Plaintiffs incorrectly contend that Facebook seeks to destroy the data and therefore has the burden under Rule 26(c) to show "that 'specific prejudice or harm will result if no protective order is granted.'" Opp. at 8 (quoting *Phillips ex rel. Ests. of Byrd v. Gen. Motors Corp.*, 307 F.3d 1206, 1210-11 (9th Cir. 2002)). But the case Plaintiffs rely on, *Phillips*, addressed a different question. It concerned a motion for a protective order against the public disclosure of information filed under seal, 307 F.3d at 1209-10, not the undue burden of producing inaccessible ESI under Rule 26(b)(2)(B). Rule 26(b)(2)(B) allows a party to move for a protective order on a showing "that the information is not reasonably accessible because of undue burden or cost." In any event, Facebook's Motion also demonstrated the substantial harm it would suffer if required to produce the Mobile and Web Tables, in whole or in part, and that such harm could not be justified by the marginal relevance of some data that might be found in the Tables.

Instead of meeting their burden of establishing good cause for the production of inaccessible data—or for that matter, addressing the fact that the Tables are inaccessible—Plaintiffs composed a "series of specific questions" falling outside the scope of the Motion that Facebook must purportedly answer before the Special Master can issue a ruling. *See* Opp. at 13-15. These questions, which read like requests for production ("Can Facebook provide documentation enumerating . . .") and interrogatories ("Is it possible to use the ████████████ information to estimate . . .") should be ignored: they are irrelevant to the issue of inaccessibility and to the relief Facebook's Motion seeks.

1. Plaintiffs have not shown that the benefit of producing data from the Mobile and Web Tables outweighs the burden on Facebook.

Plaintiffs fail to show that data from the Tables is within the scope permitted by Rule 26(b)(1)— that it is relevant and proportional considering such factors as "the importance of the discovery in resolving the issues, and whether the burden or expense of the proposed discovery outweighs its likely benefit." Fed. R. Civ. P. 26(b)(1). Nor have Plaintiffs shown that this data is within the scope of the

2

case that Judge Chhabria has allowed to proceed (because it is not).

<p style="text-align:center">a.   Plaintiffs fail to show that the relevance, importance, or benefit of data
from the Mobile and Web Tables outweighs the burden of production.</p>

To justify the heavy burden of producing data from an inaccessible data source, the requesting party must establish that there is good cause to permit the discovery under Rule 26(b)(2)(C), notwithstanding the data's inaccessibility due to undue burden or cost. Fed. R. Civ. P. 26(b)(2)(B). When evaluating the "importance of the discovery to the issues" component of good cause, courts assess whether the discovery goes to issues that "are at the very heart of [the] litigation." *Oxbow Carbon & Mins. LLC v. Union Pac. R.R. Co.*, 322 F.R.D. 1, 8 (D.D.C. 2017). Plaintiffs fail to engage with the applicable standard, instead resorting to the unfounded contention that some of the data in the Mobile and Web Tables might be relevant because it could be used as a "proxy for disclosure" to third parties. Opp. at 10-11. Contrary to Plaintiffs' speculation, neither Table establishes whether data was transferred to third parties. The Tables show which API calls were *made*, not what data was *returned*.

Even if the data from Mobile and Web Tables could theoretically be used as a proxy, Plaintiffs fail to explain how such a proxy would matter to the issues at the heart of this case. In reality, any such proxy would be immaterial for two reasons: (1) the data comes from the wrong time period and (2) the volume of data provided to third parties is not at issue in the case.

First, the Mobile and Web Tables cover a time period that has little relation to the issues in this case. As Facebook has explained, the data in the Mobile and Web Tables dates back ██████████████ and thus could have only minimal relevance to the alleged misconduct in Plaintiffs' Complaint, which stems from friend-sharing functions Facebook discontinued years before. Mot. at 11-12.

The only potentially relevant behavior Plaintiffs identify as extending ████████████ is the allegation that a handful of "whitelisted" developers had access to data of users' Facebook friends until June 2018, ████████████████████████████████████ The API calls of these few developers amount to a fraction of the data to which Plaintiffs demand access. Even that small fraction is inaccessible under Rule 26. And, in any event, calls from those developers are summarized in the Method Table (which Facebook is willing to produce), and Facebook has responded to interrogatories requesting information about the time periods in which whitelisted apps had access

<p style="text-align:center">3</p>

Gibson, Dunn &
Crutcher LLP

to friends' data. So Plaintiffs have access to the data they claim to seek. Since requiring production of Mobile and Web Tables would provide no incremental discovery benefit to Plaintiffs—and would impose substantial undue burden on Facebook—the Tables are outside the scope of discovery established by Rule 26(b)(1).

Opp. at 12. That email is irrelevant to the live theories of the case: Plaintiffs have never contended that they suffer an actionable privacy injury whenever an app that the Named Plaintiffs' friends use to organize their contacts has access to their email addresses.

      In sum, the date range of the Mobile and Web Tables establishes that they contain, at most, a negligible quantity of potentially relevant information. For this reason alone, the burden of producing such limited information outweighs the negligible (if any) benefit from its production.

      Second, the purported relevance of the data from the Mobile and Web Tables is extremely attenuated from the issues in the case. Plaintiffs offer only conclusory speculation that some of the data could serve as a "proxy" for the volume of data delivered to third-party apps. But the data reflects only what apps requested, not what was delivered in response. And the volume of data requested has no correlation to whether private user information was improperly disclosed. Even getting to the point of identifying the volume of data *requested* as a proxy for the volume of data *returned* would require Facebook to undertake the laborious process described in the Motion of restoring from cold storage and   over the course of several years. Mot. at 8-9.

      Setting aside the insurmountable practical barriers to identifying Plaintiffs' desired proxy for the volume of data disclosed to third-party apps, once in hand, that proxy would not reveal anything of importance. The *volume* of data provided to third parties is not at issue in the case; the issue is whether the user authorized the third party to receive data. The data in the Mobile and Web Tables have nothing to say about that issue: a crude approximation of the volume of data disclosed does not indicate whether a user consented to the disclosure of any particular information in that volume. Nor does it show

whether the disclosed data was private or sensitive. A rough proxy for volume of data transferred to third parties—information that is not itself probative of any alleged wrongdoing—cannot outweigh the burden on Facebook of producing data that Plaintiffs concede is not reasonably accessible.

b. Plaintiffs' proposed solution of sampling the Tables is unduly burdensome and would yield only irrelevant data.

As to burden, Plaintiffs focus on the annual expense of more than ▮▮▮▮▮ to store the more than ▮▮▮▮▮ of API Call Log data, Opp. at 9, even though Facebook's Motion solely concerns *production* of the data, not its *preservation*. Rather than contest the magnitude of the burden involved in analyzing and producing the data, Plaintiffs insist there must be some way for the data in the Tables to be "extracted, sampled or snapshotted." *Id.* at 3. As explained in Facebook's Motion and Ms. Ji's Declaration, there is not. The Special Master should reject any sampling proposal for three reasons.

First, neither the Mobile Table nor the Web Table can be sampled without significant cost, time, and burden. Facebook's Motion and Ms. Ji's Declaration both make clear that a sampling of the Mobile and Web Tables is highly impractical, as these tables are not only burdensome to search, but the data set returned is still too large to be transmitted out of Facebook. As Ms. Ji explained:

- If Facebook were to try to narrow the data within the Tables, each partition of *each* table must be separately searched for that data. The majority of partitions would first need to be restored to warm storage, adding several hours of processing time to searches. Searching each partition of each of the tables for certain values would therefore be an iterative process that would take *several years* to run across both tables. Ji Decl. ¶¶ 19, 28, 35.

- Even a targeted search of the Tables yields a tremendous amount of data. For example, Ms. Ji's searches within a single partition for ▮▮▮▮▮▮▮▮▮▮▮▮ of data in the Web Table and ▮▮▮▮▮▮ of data in the Mobile Table. *Id.* ¶¶ 26-27, 33-34.

- To be transferred out of Facebook, the data must be decompressed into flat .csv files. When decompressed, the ▮▮▮▮▮▮ data from the Web Table and ▮▮▮▮▮▮ data from the Mobile Table would fill approximately ▮ 50 GB .csv files. This process would take nearly 20 hours to complete. *Id.* ¶¶ 41-42.

- The .csv files must then be transferred to external hard drives (costing more than $1,000 each). Each transfer of a single 50 GB .csv file would take approximately 5 to 8 hours. *Id.* ¶ 43. It would take **over 30 days** to transfer the results of Ms. Ji's searches onto hard drives.

- Ms. Ji's estimates did not incorporate the delays and disruptions that would necessarily occur, including the other demands on her time, the possibility for human error, and outages and other limitations on processing capacity, or the time it would take to manage this project

and confirm the reliability of the data being transported. *Id.* ¶¶ 48-50.

Ultimately, it would take over a month, if not several months, to run a search on even a single partition of each Table and export the results to Plaintiffs. A broader search of and export from the Tables would increase the time, expense, and overall burden of this project by orders of magnitude. Thus, the sampling that Plaintiffs casually propose here is simply not feasible. *Compare BlackRock Allocation Target Shares v. Wells Fargo Bank, N.A.*, 2017 WL 953550, at *3 (S.D.N.Y. Mar. 10, 2017) (rejecting sampling that would cost "hundreds of thousands, if not millions, of dollars, will require months to conduct, and will likely result in challenges to the admissibility of evidence.").

Plaintiffs' contention that it is "likely that once Facebook provides more information about the Mobile and Web Tables, Plaintiffs could provide further guidance regarding how to generate more limited, meaningful samples from single partitions of the Mobile and Web Tables" (Opp. at 10) ignores the evidence before the Special Master and fails to appreciate the magnitude of the burden that sampling would impose on Facebook. Plaintiffs also ignore that Facebook has already provided substantial information about each of the Tables at issue, including by providing samples of the data and identifying the contents of each field. Thus, there is no "guidance" for Plaintiffs to provide upon further inquiry, as there is no "limited, meaningful" sample of these tables that can be reasonably exported outside of Facebook. None of the "questions" Plaintiffs pose change this fact.

Second, any sample produced from the Web and Mobile Tables would not be relevant to Plaintiffs' standing or any of the claims that Judge Chhabria allowed to proceed in his motion-to-dismiss order. *See Gilead Scis., Inc. v. Merck & Co., Inc.*, 2016 WL 146574, at *1 (N.D. Cal. Jan 13, 2016) (discovery must be relevant and proportional). The remaining claims Judge Chhabria allowed to proceed are friend sharing, whitelisting, specific partner relationships, and enforcement. Dkt. 298 at 6. As Facebook has explained, the Web and Mobile Tables do not contain information that would allow Plaintiffs to identify whether sensitive data was shared with a third party. Mot. at 11. They also post-date the relevant time period for Plaintiffs' core claims, including friend-sharing and whitelisting. *See Bultena v. Wash. State Dept. of Agriculture*, 2017 WL 11565179, at *1 (E.D. Wash. Nov. 22, 2017) (finding that production of documents outside the relevant time period "would be neither relevant nor proportional to the needs of the case"). Given that the Tables in their entirety do not show whether

FACEBOOK, INC.'S REPLY IN SUPPORT OF MOTION FOR A PROTECTIVE ORDER AGAINST PRODUCTION OF API CALL LOGS
CASE NO. 3:18-MD-02843-VC

Plaintiffs' data was accessed through friends' permissions, it is unfathomable that some smaller portion could be used to prove their claims. And Plaintiffs are not entitled to data for any purpose not related to the claims in this case. *Zewdu v. Citigroup Long Term Disability Plan*, 264 F.R.D. 622, 626 (N.D. Cal. 2010) ("Discovery [requests] must be narrowly tailored . . . and must not be a fishing expedition.").

The production of a sample from the Tables is particularly inappropriate in this privacy class action, where stray pieces of data will not be probative to establishing class wide liability. In a class action, statistical or sampled evidence can only be used to "prov[e] classwide liability" where "each class member could have relied on that sample to establish liability if he or she had brought an individual action." *Tyson Foods, Inc. v. Bouaphakeo*, 136 S.Ct. 1036, 1040 (2016); *see Desilva v. N. Shore-Long Island Jewish Health Sys., Inc.*, 27 F. Supp. 3d 313, 331 (E.D.N.Y. 2014). The data in the Tables cannot be used to prove the Named Plaintiffs' individual claims. Not only does the data post-date friend-sharing, but it also does not show whether data was actually transferred to a third party, much less whether the transfer was to an app the user downloaded. Because this data has no bearing on the Named Plaintiffs' individual claims, a sample cannot be used to prove class-wide claims.

Third, Facebook's willingness to produce the summary data in the Method Table obviates any need for sampling. The only purpose of sampling would be to determine whether there is good cause for production of the Web and Mobile Tables despite their inaccessibility. The data in the Method Table—a considerably less burdensome source that provides all of the information Plaintiffs seek— makes it impossible for Plaintiffs to make this good-cause showing. Again, Rule 26 provides that courts "*must* limit the frequency or extent of discovery otherwise allowed" if the discovery sought "can be obtained from some other source that is more convenient, less burdensome, or less expensive." Fed. R. Civ. P. 26(b)(2)(C) (emphasis added). Production from the Method Table will give Plaintiffs access to the information about volume of API calls that Plaintiffs say they intend to use as a proxy for the volume of data transferred. *See* Ji Decl. ¶¶ 53-56, Ex. E. Rule 26 does not require Facebook to *also* bear the burden and cost of producing the same data from the inaccessible Web and Mobile Tables.[1]

---

[1] While Facebook's motion is concerned with the burden and expense of production—not preservation—Plaintiffs are wrong to trivialize the burden of preserving the data, which is substantial. Plaintiffs point out that Facebook has 18 data centers capable of storing 1,000 petabytes of data in cold storage and thus suggest the additional burden of storing ▮▮▮▮▮▮▮ of data is insignificant.

7

1

2

3

4

5

c. To the extent that Facebook is ordered to produce any portion of the Tables, Plaintiffs should share in the significant costs.

If the Special Master is inclined to order even partial production of the Web and Mobile Tables, the Special Master should shift the costs of any such production to Plaintiffs. As Plaintiffs apparently believe that the costs at issue are negligible, they should take no issue with paying them.

6

7

8

9

10

11

12

13

14

15

Courts have discretion to shift all or part of the costs of production to the requesting party. *Oppenheimer Fund, Inc. v. Sanders*, 437 U.S. 340, 358 (1978). This is particularly true where a party seeks production of inaccessible data sources. *See Zubulake v. UBS Warburg LLC*, 216 F.R.D. 280, 284 (S.D.N.Y. 2003). Courts have also held that cost shifting is appropriate *even in instances where data is accessible* as part of enforcing proportionality limits. *See F.D.I.C. v. Brudnicki*, 291 F.R.D. 669, 674 (N.D. Fla. 2013). Costs of production may also be shifted to the requesting party based on, among other factors, the likelihood of discovering critical information, the availability of such information from other sources, the relative benefit to the parties of obtaining the information, and the total cost associated with the production. *See Rowe Entertainment, Inc. v. The William Morris Agency, Inc.*, 205 F.R.D. 421, 428-29 (S.D.N.Y. 2002).

16

17

18

19

20

21

22

23

24

Here, Facebook has demonstrated that the Mobile and Web Tables are inaccessible and the costs of producing even a portion of the data they contain would be highly disproportionate to the needs of the case. While Facebook's Motion did not focus on the tangible costs of production—including employee time spent, the costs of warm storage (which is more expensive than cold storage), and the costs of dozens of hard drives—they would be substantial.[2] Courts have shifted costs to the requesting party in cases involving far lighter production burdens, even where the producing party was a large company. *See U.S. ex rel. Carter v. Bridgepoint Educ., Inc.*, 305 F.R.D. 225, 243 (S.D. Cal. 2015) ("Plaintiffs will have much ESI to sift through in the months ahead. . . it is fair they pay the price of its production in the . . . form they deem so invaluable for the prosecution of their own case."); *Zubulake*,

25

26

27

28

But Facebook uses these centers to provide services to more than *two billion* users. Spending ▮ ▮ per year and using ▮▮▮▮▮▮▮▮▮▮▮▮▮▮ for one irrelevant discovery request in one civil litigation is a waste of valuable resources. If that level of expense is proportional to the needs of this case, the proportionality requirement is toothless.

[2] The Special Master has directed the parties not to submit new evidence on Reply. Facebook will provide the Special Master with greater detail regarding the costs of production upon request.

8

FACEBOOK, INC.'S REPLY IN SUPPORT OF MOTION FOR A PROTECTIVE ORDER AGAINST PRODUCTION OF API CALL LOGS
CASE NO. 3:18-MD-02843-VC

Gibson, Dunn & Crutcher LLP

216 F.R.D. at 289 (shifting 25% of the $166,000 cost of inaccessible data to the requesting party when the data was relevant but not "indispensable"); *Wiginton v. CB Richard Ellis, Inc.*, 229 F.R.D. 568, 575 (N.D. Ill. 2004) (shifting 75% of the cost to requesting party given the low relevance and costs amounting to several hundred thousand dollars); *OpenTV v. Liberate Techs.*, 219 F.R.D. 474, 479 (N.D. Cal. 2003) (shifting 50% of the cost to requesting party given the undue burden and expense, including 150 hours of employee work, involved in extracting and copying source code"). Facebook has also demonstrated Plaintiffs will not find "critical information" in the Web and Mobile Tables, much less that the "relative benefits" to Plaintiffs could come close to outweighing the vast time and costs associated with producing the data. And Facebook has presented a reasonable alternative.

Plaintiffs should not be permitted to insist upon burdensome discovery from the exact source of their choosing without sharing in that burden. If Facebook is ordered to make a production of this data, the Special Master should allow Facebook to submit evidence of the costs associated with that production and order that Plaintiffs share in the expense of any production they insist on.

2. Plaintiffs have not shown that the Method Table is an inadequate alternative to the Mobile and Web Tables.

It is not Facebook's burden to show that the discovery Plaintiffs seek can be obtained from a less burdensome source; it is Plaintiffs' burden to establish good cause for the production of the inaccessible information, and they have failed to do so. *See* Mot. at 6-7, 10-13. Nonetheless, Facebook has explained that by November 30, 2021, it can produce a manageable data set from the Method Table that summarizes the ████████████████ and contains data from the relevant time (unlike the Mobile and Web Tables). Mot. at 13. The Ji Declaration describes the contents of the Method Table in detail and attaches a list of the data fields it contains. Ji Decl. ¶¶ 53-56.[3] These include fields ██. *Id.* ¶ 55, Ex. E. Plaintiffs have no response beyond a conclusory contention that the Ji Declaration provides "too few details to understand [the Method Table's] contents," and immaterial complaints that Facebook did not previously disclose the Method Table. Opp. at 3, 4, 8. Critically, Plaintiffs do not explain how the

---

[3] To comply with the Special Master's request that citations to exhibits be hyperlinked, Facebook has re-attached Ms. Ji's Declaration to this brief in order to link to it. No changes have been made to the Ji Declaration or its Exhibits since it was initially submitted.

Method Table is insufficient to meet their alleged need for data to use as a "proxy" for data transfer.

Given the Method Table's availability, Plaintiffs cannot possibly establish good cause for requiring Facebook to incur the cost and burden of producing data from the Mobile and Web Tables. Rule 26 instructs that the Court "*must* limit the extent of discovery" where "the discovery sought … can be obtained from some other source that is more convenient, less burdensome, or less expensive." Fed. R. Civ. P. 26(b)(2)(C)(i) (emphasis added). The Special Master should equally do so here.

B.   Plaintiffs present an inaccurate history of the parties' negotiations.

Plaintiffs pad their brief with accusations that Facebook has "sandbagged" them with its Motion and concealed relevant information from them for "years." Because Plaintiffs' accusations are untrue, Facebook corrects them below. But none of those accusations is relevant to Facebook's Motion.

- *Facebook disclosed the tables in 2018*. Plaintiffs say Facebook "only disclosed the existence of any API call logs in July" 2021. Opp. 4, 8. But, on November 5, 2018, Facebook informed Plaintiffs the "web" and "mobile" API call logs contained records of API calls and explained "the overwhelming majority of [data in these tables] would be irrelevant" and "retaining API call logs in perpetuity [is] . . . impractical."[4]

- *Facebook urged Plaintiffs to narrow their requests*. During the parties' meet and confers regarding API call data, Facebook made clear the data contained in the logs is irrelevant because, among other things, they do not reflect whether data was accessed by third parties and are not intended for human consumption, and the logs were tremendously burdensome to produce. Facebook tried to work with Plaintiffs to tailor their requests and Plaintiffs never insisted they needed vast amounts of technical data to hold these negotiations. Because the parties were unable to reach resolution, Facebook subsequently raised the API call logs in mediation. The mediators declared impasse on the issue before Facebook was able to prepare a proposal regarding an alternative production from the Method Table.

- *Facebook has produced sufficient technical information*. Facebook has endeavored to provide Plaintiffs with the technical information they seek, including by producing historical versions of the developer manual, training videos about the social graph, articles about its data infrastructure, internal Wikis relating to Facebook's data storage systems and search capabilities, and more. Regardless, Plaintiffs' hunger for ever more information about Facebook's data infrastructure has no bearing on Facebook's Motion.

II.   CONCLUSION

Facebook respectfully requests that the Special Master grant its motion for a protective order holding that Facebook is not required to produce data from the Mobile and Web Tables in this litigation.

---

[4] The Special Master has directed the parties not to submit new evidence on Reply. Facebook will, of course, provide the Special Master with a copy of this letter upon request.

FACEBOOK, INC.'S REPLY IN SUPPORT OF MOTION FOR A PROTECTIVE ORDER AGAINST PRODUCTION OF API CALL LOGS
CASE NO. 3:18-MD-02843-VC

Gibson, Dunn & Crutcher LLP

1   Dated: November 2, 2021   **GIBSON, DUNN & CRUTCHER, LLP**

2   By: _/s/ Russ Falconer_
                                             Orin Snyder (*pro hac vice*)
3   osnyder@gibsondunn.com
                                             200 Park Avenue
4   New York, NY 10166-0193
                                             Telephone: 212.351.4000
5   Facsimile: 212.351.4035

6
                                             Deborah Stein (SBN 224570)
7   dstein@gibsondunn.com
                                             333 South Grand Avenue
8   Los Angeles, CA 90071-3197
                                             Telephone: 213.229.7000
9   Facsimile: 213.229.7520

10
                                             Joshua S. Lipshutz (SBN 242557)
11   jlipshutz@gibsondunn.com
                                             1050 Connecticut Avenue, N.W.
12   Washington, DC 20036-5306
                                             Telephone: 202.955.8500
13   Facsimile: 202.467.0539

14
                                             Kristin A. Linsley (SBN 154148)
15   klinsley@gibsondunn.com
                                             Martie Kutscher (SBN 302650)
16   mkutscherclark@gibsondunn.com
                                             555 Mission Street, Suite 3000
17   San Francisco, CA 94105-0921
                                             Telephone: 415.393.8200
18   Facsimile: 415.393.8306

19
                                             Russ Falconer (*pro hac vice*)
20   rfalconer@gibsondunn.com
                                             2001 Ross Avenue, Suite 2100
21   Dallas, TX 75201-2911
                                             Telephone: 214.698.3100
22   Facsimile: 214.571.2900

23
                                             Attorneys for Defendant Facebook, Inc.
24

25

26

27

28

FACEBOOK, INC.'S REPLY IN SUPPORT OF MOTION FOR A PROTECTIVE ORDER AGAINST PRODUCTION
OF API CALL LOGS
CASE NO. 3:18-MD-02843-VC

Gibson, Dunn &
Crutcher LLP

GIBSON, DUNN & CRUTCHER LLP
Orin Snyder (*pro hac vice*)
 osnyder@gibsondunn.com
200 Park Avenue
New York, NY 10166-0193
Telephone: 212.351.4000
Facsimile: 212.351.4035

Kristin A. Linsley (SBN 154148)
 klinsley@gibsondunn.com
Martie Kutscher (SBN 302650)
 mkutscherclark@gibsondunn.com
555 Mission Street, Suite 3000
San Francisco, CA 94105-0921
Telephone: 415.393.8200
Facsimile: 415.393.8306

GIBSON, DUNN & CRUTCHER LLP
Deborah Stein (SBN 224570)
 dstein@gibsondunn.com
333 South Grand Avenue
Los Angeles, CA 90071-3197
Telephone: 213.229.7000
Facsimile: 213.229.7520

Joshua S. Lipshutz (SBN 242557)
 jlipshutz@gibsondunn.com
1050 Connecticut Avenue, N.W.
Washington, DC 20036-5306
Telephone: 202.955.8500
Facsimile: 202.467.0539

Attorneys for Defendant Facebook, Inc.

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

IN RE: FACEBOOK, INC. CONSUMER PRIVACY USER PROFILE LITIGATION, This document relates to: ALL ACTIONS	CASE NO. 3:18-MD-02843-VC **DECLARATION OF MENGGE JI IN SUPPORT OF FACEBOOK, INC.'S MOTION FOR A PROTECTIVE ORDER AGAINST PRODUCTION OF API CALL LOGS** Discovery Special Master Daniel Garrie, Esq.

DECLARATION OF MENGGE JI IN SUPPORT OF
FACEBOOK, INC.'S MOTION FOR A PROTECTIVE ORDER AGAINST
PRODUCTION OF API CALL LOGS

HIGHLY CONFIDENTIAL

I, Mengge Ji, declare as follows:

1.   I am a Data Scientist at Facebook, Inc. ("Facebook"). My job responsibilities include, among other things, understanding and working with Facebook's data systems, writing queries and conducting analyses of these data, researching Facebook's data and related technologies, and locating, analyzing, and exporting data for production in litigation and other legal matters. I submit this declaration in support of Facebook's motion for a protective order as against the production of certain of Facebook's API call logs. Unless otherwise stated, I have personal knowledge of the facts set forth herein, and, if called as a witness, I could and would competently testify thereto.

2.   My understanding is that Plaintiffs in this matter have requested that Facebook produce to them all "logs of API Calls made by Third Parties for the Content and Information of Friends of Installing Users," including specific information about, among other things, "the response status of those Calls (*e.g.*, success or reason for failure)" and "[t]he Content and Information delivered to Third Parties on each API in response to these Calls."

3.   I also understand that Plaintiffs have issued additional requests for information relating to the number of API calls made by certain "Third Parties" and "Business Partner[s]," including "the number of calls . . . made to each such API . . . each month" and "the volume of data transferred from each such API . . . each month."

1.   Background on Facebook's APIs and Call Logs

4.   Application Program Interfaces ("APIs") are standard computing protocols used throughout the digital world. Each time a user visits an app, the user necessarily interacts through an API. And every company that allows third parties to communicate with their servers

2   **HIGHLY CONFIDENTIAL**

uses APIs. As just one example, email programs like Outlook or Mail on iOS communicate with

the user's email provider (Yahoo, Gmail, etc.) through APIs to obtain the user's messages.

5.   Facebook has a series of APIs that allow for data to be transferred to and from

Facebook's Platform. The Facebook Platform is powered by a series of databases that Facebook

relies upon to host and run the Facebook products and which work in tandem to provide

Facebook users a seamless experience as they access different categories and types of content

posted on the Platform, by themselves or other users. APIs facilitate the transfer of data from the

databases that support the Platform to websites and applications, including Facebook's own,

proprietary mobile apps. For instance, if a Facebook user pulls up her profile photo via the

Facebook mobile app, the Facebook mobile app will "call" a Facebook API to request that photo

to be displayed locally on the user's phone.

6.   Third party mobile apps and websites can also connect to Facebook through its

APIs. For instance, if a user chooses to log in to NYTimes.com with his Facebook account

instead of creating an account for that site, NYTimes.com will call the Facebook APIs to verify

the user's identity. Facebook has a series of public APIs that are published on its developer

website (https://developers.facebook.com/), and app developers can use them or request to use

them to send or receive information from the Facebook Platform.

7.   Facebook maintains internal logs that contain the number of API calls made to the

Facebook Platform. Two of the data tables that contain ███████████████   Facebook

maintains regarding API call activity are the ███████████ table and the ███████████ table.

2.   **The ███████████ and ███████████ Tables**

8.   The ███████████ table contains logs of API calls ███████████████

██

HIGHLY CONFIDENTIAL

██

██   The data in

the ████████████ table dates back to ████████████ As of October 6, 2021, Facebook had

████████████ data in this table.

     9.   The categories of data logged in the ████████████ table include, among other

things, ██

███. A true and

correct list of the data fields reflected in the ████████████ table is attached hereto as **Exhibit**

A. A true and correct sample of ten lines of data from the ████████████ table, with the user

identifiable information redacted, is attached hereto as **Exhibit B**.

     10.   The ████████████ table contains logs of API calls ████████████████████

██

███. The

data in the ████████████ table dates back to ████████████ As of October 6, 2021, Facebook

had ████████████ data in this table.

     11.   The categories of data logged in the ████████████ table include, among other

things, ██

████████████████████████████████. A true and correct list of the data fields

reflected in the ████████████ table is attached hereto as **Exhibit C**. A true and correct sample

of ████████████ from the ████████████ table, with the ████████████████████ redacted,

is attached hereto as **Exhibit D**.

     12.   Several of the fields in each of these tables do not contain ████████████

██

4   **HIGHLY CONFIDENTIAL**

██████████████████████████. This is an extra step of analysis that would need to be conducted if these tables were ever to be used or analyzed outside of Facebook's environment.

13.   The ██████████ and ██████████ tables do not contain data demonstrating whether data was returned in response to an API call, what the data that was returned may have been (if it was returned at all), or the volume of data that may have been returned (if it was transferred at all).

14.   I understand that there are many reasons why data would not be returned in response to an API call. These reasons include, but are not limited to, ██████████████ ██ ██ ██

A.   Storage of the ██████████ and ██████████ tables

15.   The ██████████ and ██████████ tables are stored in a highly compressed state in Facebook's internal decentralized, structured data warehouse, known as "Hive." Hive is designed to ingest and process data for internal analysis, not to export data into files that can be transferred outside of Facebook. As discussed further below, accessing and analyzing large data tables in Hive is a multi-step process that can take a considerable amount of time, depending on the volume of data being accessed and the complexity of the analysis or querying needed.

16.   The active ██████████ and ██████████ tables are currently maintained pursuant to ███ business retention periods ██████████████████████ ██ ██

HIGHLY CONFIDENTIAL

██

████████████████████████████████████

17.   ██

██

██████████████████████████████████████

18.   Tables in cold storage cannot be reviewed or analyzed while they remain in cold

storage. If Facebook wants to conduct any analysis of data in cold storage, a Facebook data

scientist must first restore it from cold storage to an accessible location, referred to as "warm"

storage. It takes approximately 1 day of processing to restore ███████████ of cold storage

data. Restored cold storage data is only made available to legal team members.

19.   The ████████ and ████████ tables are stored in "partitions" or

segments. These tables are partitioned ███████████████████████████████████████

████████   Because each ███ is partitioned off from the others, Facebook cannot quickly or

efficiently run searches across multiple ██████████████████   Instead, Facebook's

system is optimized for running searches within an individual partition. As a result, Facebook

cannot search these entire tables for a specific value, for instance ███████████ while in

cold storage, and instead must restore individual partitions to warm storage and then search the

partitions individually. Due to the same limitations on accessing and searching multiple

partitions of data, Facebook also cannot filter these tables for ████████████████████

████████████████████████ without first restoring partitions from cold

storage piecemeal.

20.   Accordingly, to search or analyze the data in either the ███████████ table or

██████████ tables, a Facebook data scientist must first restore a partition of the data to warm

HIGHLY CONFIDENTIAL

storage. Due to limitations on computing and server capacity, only a limited number of partitions can be restored and analyzed at any given time.

21.   The ███████████ and ███████████ tables are very large data tables, even in comparison to other data sets at Facebook and even when compressed in Facebook's specialized data storage environment. The volume of the ███████████ ███████████████████ in its most highly-compressed form in cold storage. The volume of the ███████████ table is currently █████ in its most highly-compressed form. Maintaining the high volume of this data ███████████ is beginning to tax Facebook's data storage system, which is one of the largest and most sophisticated data storage systems in the world. For instance, I understand that it is costing Facebook approximately ████████████████████████████████████ data ███████████

██████

B.   Sampling of the ███████████ table

22.   In order to assess how long it might take to produce data from the ███████████ table to the Plaintiffs in this action, I recently restored and analyzed one partition of data from the ███████████ table dated ███████████

23.   The volume of the ███████████ partition is ███████████ in cold storage. For reference, 1 TB is four-times the size of a standard laptop computer hard drive of 256 GB, so the *compressed* volume of the data for this █████ equals ████████████████████████████. The size of the data in each partition in the ███████████ table varies ███████████████ ███████████████████ For example, the most recent partitions of the table in cold storage are around ███████████

24.   In order to analyze the ███████████ data, I first restored this partition to warm storage. The length of time it takes to restore data from cold storage depends on how many

HIGHLY CONFIDENTIAL

requests are pending at the same time and the computing power available, which varies

depending on the time of day. In this instance, it took approximately 2 hours of processing time

to restore this single partition into warm storage. Based on the volume of this partition and

others in this table, I estimate that Facebook presently has the computing and server capacity to

host up to five partitions of this table in warm storage at a time. As a result, Facebook can

analyze no more than five partitions of this table at a given time in the ordinary course. In other

words, an effort to simultaneously analyze more than five partitions of this data would displace

and disrupt all other efforts by the legal team to perform other data analyses for this and other

matters, including for the ███████████ table.

25.   The ██████████ partition of the ████████████ table contains

███████████████████ information. As a point of reference, a Microsoft Excel file can only

hold up to 1 million rows of data. If translated into Excel, then, this one partition of the

███████████ table would comprise more than ██████████ Excel files.

26.   As I mentioned in paragraph 8 above, this table only contains logging for API

calls ██

███

The partition only contained data for ████████████. It took 1 hour and 14 minutes for this

search to complete.

27.   ██

███

███

███

████████████

HIGHLY CONFIDENTIAL

28.   Based on the results of this exercise, I estimate it would take 2,378 hours (█████

███████ in cold storage) to restore each of the partitions in the ████████ table to

warm storage and additional 1,279 hours (████████████) to run a single search on each.

Assuming I worked exclusively on this task for 8-hours per day, 260 days each year, it would

take nearly two years for me to run this same search on ████████ of this data (through the

████████ partition). This is also a conservative estimate, ████████████████

████████████████████. This estimate also does not account

for the other demands on Facebook's personnel and resources (including the need to use these

same systems to respond to requests for other legal matters), occasional outages, or many other

variables.

C.   Sampling of the ████████ table

29.   I also restored and recently analyzed one partition of data from the ████████

table dated ████████

30.   The volume of the ████████ partition is 19 TB in cold storage. The size of

the data in each partition in the ████████ table varies ████████████████

████ For example, the most recent partitions of the table in cold storage are around ████████.

31.   In order to analyze the ████████, I first restored this partition to warm

storage. The length of time it takes to restore data from cold storage depends on how many

requests are pending at the same time and the computing power available, which varies

depending on the time of day. In this instance, it took approximately ████ of processing

time to restore this partition into warm storage for analysis. Based on the volume of this

partition and others in this table, I estimate that Facebook presently has the computing and server

9   **HIGHLY CONFIDENTIAL**

capacity to host 20 partitions of this table in warm storage at a time. As a result, Facebook can

only analyze 20 partitions of this table at a given time in the ordinary course.

32.   The ██████████ partition of the ██████████ table contains

██████████████████████ information. If exported to Excel, then, this partition of the

██████████ table would alone comprise ███████ Excel files.

33.   I ran a search for ███

██

It took 43 minutes for this search to run.

34.   Of the ██████████████ of data in this table, ████████████████████ were

returned for these ██████████ ██

██████████

35.   Based on the results of this exercise, I estimate it would take 2,970 hours (2.5

hours per partition in cold storage) to restore each of the partitions in the ██████████ table to

warm storage and an additional 935 hours (45 minutes per partition) to run a single search on

each. Assuming I worked exclusively on this task for 8-hours per day, 260 days each year, it

would take nearly two years for me to run this same search on each partition of this data (through

the ██████████ partition). This is also a conservative estimate, ██████████████████████

██ This estimate also does not account

for the other demands on Facebook's personnel and resources, occasional outages, or many other

variables. This estimate is also *in addition to* the two years it would take to search the

██████████████ table.

HIGHLY CONFIDENTIAL

36.   As a result, it is neither feasible nor practicable to restore all of the partitions of

the ███████████ and/or ███████████ tables at the same time and ███████████

███.

3.   The Burden of Production of API Call Log Data

37.   As noted, the ███████████ and ███████████ tables are stored in Facebook's

internal structured data warehouse, known as "Hive." This data warehouse is not optimized to

export data in a manner that can be transferred outside of Facebook. For Facebook to export and

produce data from Hive tables, the data must be decompressed, broken into small pieces, and re-

written to .csv files (which can be opened with Excel). This would be a manual, iterative process

conducted by Facebook data scientists on the legal team. As a result, the timeline for completing

an export needs to account not only for the limitations on Facebook's computing capacity, the

size limitations on our mechanisms for export, and the availability of Facebook's technical

resources, but also the possibility of human error, which necessitates real-time quality checks of

the data downloads.

38.   Before any analysis or export can be conducted, the data must be transitioned to

warm storage in order to be reviewed, analyzed, or exported. As discussed above in Paragraphs

28 and 35, I estimate it would take 5,348 hours to restore all of the partitions of ███████████

and ███████████ tables in to warm storage, which is more than 2.5 years of work, assuming I

work exclusively on this task for 8 hours per day, 260 days each year, and do not run any

searches to narrow the data set. Running searches on the data set would take at least an

additional year.

11   **HIGHLY CONFIDENTIAL**

39.   After data has been restored into warm storage, it is necessary to decompress the

data, which requires that Facebook take several additional steps, which are themselves time-

consuming.

40.   First, a Facebook data scientist must write entirely new code to query the data and

demarcate each partition into chunks of data that can be effectively transported to and/or

received by the recipient. For instance, if the recipient intends to use Excel to analyze the data,

each partition must be broken into chunks of 1 million rows each, which is the maximum number

of rows that can be opened in an Excel file. When Facebook must produce Hive data in

litigation, it typically produces it in Excel form. If the intended recipient—in this instance,

Plaintiffs—had access to or intended to use specialized software to access the data (*e.g.*, SQL, R,

Stata, SAS) and had sufficient computing power to access larger files, the limiting factor would

be a combination of size of the drive used to transport the data and the maximum number of rows

that could be reliably written into a .csv and reviewed for quality control. I understand from my

colleagues at Facebook that the largest .csv files that we have handled and transferred reliably is

50 GB, which is roughly ███

████████████████████████████████ .

41.   Second, the code is run to export that data to a server, which would involve

querying each chunk of ███████████████████████ the data warehouse and then writing the

results of that query as .csv files in a decompressed flat-file format. The speed at which

Facebook can export the data from the warehouse to the server varies depending on usage of its

systems in the ordinary course of business, but based on testing and prior exports, Facebook's

current best estimate is that it takes about 40 minutes to write a 50 GB .csv file. At this rate, it

would take approximately 15 hours to extract the ██████████████ data I identified in my

HIGHLY CONFIDENTIAL

search of the ███████ partition of the ██████████ table ███████████████████████ ███████, discussed in Paragraph 27 above, to .csv format. The resulting .csv files would be approximately ██████████, or approximately ████████████████████████████████ ████ It would then take our eDiscovery and Data Science teams about 3 minutes <u>per file</u> to inspect the data to confirm that it was not corrupted or truncated during export. If we were to export the ████████████████ data I identified in my search of the █████████ partition of the █████████ table—██ ████████████—████████████████████████████████, it would take approximately 1.5 hours to inspect these files. It would therefore take approximately 16.5 hours to extract and inspect .csv files containing ██.

42.   It would separately take 2 hours to extract the ████████████████ data I identified in my search of ██ discussed in Paragraph 34 above, to .csv format. The resulting .csv files would be approximately ██████████. It would then take our e-Discovery and Data Science teams about 3 minutes <u>per file</u> to inspect the data to confirm that it was not corrupted or truncated during export. If we were to export the ████████████████ data I identified in my search of the █████████ partition of the ██ ████████████████████████████████████ each, it would take approximately 12 minutes to inspect these files. It would therefore take an additional 2 to 3 hours to extract and inspect .csv files containing ████████████████████████ ████████████████.

43.   Third, once the server is full, a data scientist must empty the server by transferring the .csv files into a local encrypted hard drive or a secure transfer site. Each 50 GB .csv file

would require 5 to 8 hours to be transferred from the Facebook servers to a physical hard drive, depending on network speeds and other variables.

44.   Decompression of the data, which is required for export of this data, results in a significant expansion in the size of the data. Within Facebook's systems, one partition of the ███████████ table comprises approximately ████████ of data and one partition of the ███████████ table comprises approximately ████████ of data. It is difficult to estimate the compression ratio before the data is actually exported. Based on testing and prior exports, Facebook's current best estimate as to how big a single partition of each of these tables would be when fully decompressed to .csv format for export is around ████████ per partition for the ███████████ table and ███████████ per partition for the ███████████ table.

45.   ██ ██ As a result, the second and third steps of this process would have to be run a minimum of 19,250 separate times for each partition of the ███████████ table. Similarly, the second and third steps of this process would have to be run a minimum of 4,750 separate times for each partition of the ███████████ table. It is not possible to expand the number of available servers without investing significantly in additional resources or interrupting other business and legal functions.

46.   The decompressed volume of a single partition of data from each of these tables—███████████████████████ table and █████████ partition for the ███████████ table—is significantly larger than any hard drives or secure transfer site that Facebook has access to. Facebook's current best estimate is that exporting a single partition of the ███████████ table would require it to load .csv files onto ███████████ encrypted hard drives that would cost approximately $1,200 each, presuming Facebook can acquire a

14   **HIGHLY CONFIDENTIAL**

sufficient number of them. Facebook's current best estimate is that exporting a single partition of the ████████ table would require it to load .csv files onto ████████ encrypted hard drives that would cost approximately $1,200 each, again presuming Facebook can acquire a sufficient number of them.

47.   The process must be repeated over and over until the export is complete. Facebook estimates that it would take 8,000 days and $70,000 in hardware costs to export the smallest, single partition of the ████████ table and 2,000 days and $17,000 in hardware costs to export a single partition of the ████████ table.[1] Replicating this process for more than ████████ of each table would multiply those timing estimates proportionally.

48.   This manual, iterative process introduces the possibility of human error, which necessitates real-time quality checks of the data downloads that take additional time and create the potential for delay.

49.   Facebook also estimates that it will take several days for senior members of the eDiscovery data scientist team conducting the export to prepare for and manage the coding and export project. After the initial exported files have been created, Facebook estimates it would take an additional month for a review team to validate the queries and data and perform additional quality control, or longer if there are issues that require remediation.

A.   **Impact on Facebook's E-Discovery Team's Resources**

50.   In total, Facebook currently estimates that, even without any other responsibilities or deadlines, it would be substantial, multi-year project to attempt to export the ████████ table and the ████████ table through the process described above. The Facebook E-Discovery team, which is responsible for assisting Facebook in-house and outside counsel in

---

[1] The shipping costs for the hard drives would add additional costs onto this project.

HIGHLY CONFIDENTIAL

active litigations and other legal matters, in addition to building and maintaining internal

infrastructure crucial to the management and preservation of data on legal hold, does not have

the time and resources required to access, analyze, and export the data in the ▓▓▓▓▓▓▓

table and the ▓▓▓▓▓▓▓▓▓ table in the manner described above.

51.   Facebook cannot materially shorten this timeline by hiring new employees

because it would take time and resources to interview, select, and onboard new employees, and

any new hires would have to be sufficiently trained regarding Facebook's systems, policies, and

procedures, which is a mandatory process that takes three weeks. The API call log data export in

particular requires specialized knowledge about Facebook's systems because the system is not

built for the purpose of exporting the extremely large amounts of data at issue—knowledge that

non-Facebook personnel do not have, given the nature of the proprietary systems and databases

employed by Facebook. For the same reason, Facebook cannot simply engage third party

consultants or temporary employees to handle this data export. Nor would adding more

servers—which would require diverting them from their use in the ordinary course of business—

necessarily reduce the estimated timeline in a linear fashion, as the export still requires

downloading the data from the server, and the manual process of writing the code and

monitoring the export. All of these options—hiring new employees, hiring contractors, and

adding servers—would also be extremely costly.

4.   **Alternative Production Proposal**

52.   I understand that Facebook has proposed to produce summary data relating to API

calls to Plaintiffs in this matter instead of data from the ▓▓▓▓▓▓▓▓ and ▓▓▓▓▓▓▓▓

tables. ▓▓

HIGHLY CONFIDENTIAL

██

███████████████████████████

53.   Specifically, I understand that Facebook has proposed to produce data from the

████████████████████ table. The ████████████████ table contains summaries of data

██ . █████████

██

██

██

██

██████

54.   Because the █████████████████████ table summarizes ██████████████, it is

substantially less burdensome to store and analyze.

55.   The ██████████████████ table contains summary data regarding API calls

made to the Facebook Platform, ██

████████ The fields contained in this table include, among other things, ████████

██

██████████████████████████████████ A true and correct list of the data fields

reflected in the ████████████████ table is attached hereto as **Exhibit E**. A true and

correct sample of ████████████ from the ████████████████ table is attached hereto as

Exhibit F.

56.   The data in the ████████████████ table dates back to ████████████

57.   The ██████████████████ table is maintained in warm storage in the Hive

data warehouse. It is currently █████ in this compressed form. A single partition ████████

HIGHLY CONFIDENTIAL

of this table is around ▮▮▮▮ in compressed form and ▮▮▮▮ in raw (*i.e.*, not compressed) form.

58.   It is difficult to estimate the compression ratio before the data is actually exported. Based on testing and prior exports, Facebook's current best estimate as to how big an export of data from this table dating from ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮ would be when fully decompressed to .csv format for export is ▮▮▮▮

59.   Facebook is currently preparing an export of data from this table dating from ▮▮▮▮▮▮▮▮▮▮▮▮▮▮ for production to Plaintiffs in this matter and estimates it will be able to produce this data by November 30, 2021.

60.   A brief summary of the three tables discussed herein is provided below, along with relevant data points regarding the burden of production.

	▮▮▮▮▮	▮▮▮▮▮	▮▮▮▮▮▮
Earliest Date	▮▮▮▮	▮▮▮▮	▮▮▮▮▮
Compressed Size	▮▮▮	▮▮▮	▮▮▮
Est. Decompressed Partition Size	▮▮▮	▮▮▮▮	▮▮▮
Est. Decompressed Size of Data Through ▮▮▮	▮▮▮	▮▮▮	▮▮▮

I declare under penalty of perjury that the foregoing is true and correct, and that I executed this Declaration on October 18, 2021, in Sausalito, California.

Mengge Ji

HIGHLY CONFIDENTIAL

Exhibit A

to Ji Declaration

Highly Confidential

REDACTED

Exhibit B

to Ji Declaration

Highly Confidential

REDACTED

REDACTED

REDACTED

REDACTED

REDACTED

REDACTED

REDACTED

Exhibit C

to Ji Declaration

Highly Confidential

REDACTED

Exhibit D

to Ji Declaration

Highly Confidential

REDACTED

REDACTED

REDACTED

REDACTED

REDACTED

Exhibit E

to Ji Declaration

Highly Confidential

REDACTED

Exhibit F

to Ji Declaration

Highly Confidential

REDACTED

PROOF OF SERVICE BY E-Mail

Re: In re: Facebook, Inc. Consumer Privacy User Profile Litigation (Special Master)
Reference No. 1200058674

I, Anne Lieu, not a party to the within action, hereby declare that on November 8, 2021, I served the attached ORDER RE: FACEBOOK'S MOTION FOR PROTECTIVE ORDER AGAINST PRODUCTION OF API CALL LOGS on the parties in the within action by electronic mail at El Monte, CALIFORNIA, addressed as follows:

Martie P. Kutscher Clark Esq.
Gibson Dunn & Crutcher
1881 Page Mill Rd.
Palo Alto, CA   94304-1125
Phone: 650-849-5300
mkutscher@gibsondunn.com
    Parties Represented:
    Facebook, Inc.

Deborah L. Stein Esq.
Alexander P. Swanson Esq.
Gibson Dunn & Crutcher
333 S. Grand Ave.
52nd Floor
Los Angeles, CA   90071-3197
Phone: 213-229-7000
dstein@gibsondunn.com
aswanson@gibsondunn.com
        Parties Represented:
        Facebook, Inc.

David J. Ko Esq.
Derek W. Loeser Esq.
Cari C. Laufenberg Esq.
Keller Rohrback LLP
1201 Third Ave.
Suite 3200
Seattle, WA   98101-3052
Phone: 206-623-1900
dko@kellerrohrback.com
dloeser@kellerrohrback.com
claufenberg@kellerrohrback.com
    Parties Represented:
    Bridgett Burk
    Cheryl Senko
    Jason Ariciu
    Jordan O'Hara
    Rev. Anthony Bell
    Samuel Armstrong
    Steven Akins
    Terry Fischer

Chris Springer Esq.
Keller Rohrback LLP
801 Garden St.
Suite 301
Santa Barbara, CA   93101
Phone: 805-456-1496
cspringer@kellerrohrback.com
    Parties Represented:
    Bridgett Burk
    Cheryl Senko
    Jason Ariciu
    Jordan O'Hara
    Rev. Anthony Bell
    Samuel Armstrong
    Steven Akins
    Terry Fischer
    Tyler King

Tyler King

Lesley E. Weaver Esq.
Ms. Anne Davis
Matthew Montgomery Esq.
Bleichmar Fonti & Auld LLP
555 12th St.
Suite 1600
Oakland, CA   94607-3616
Phone: 415-445-4003
lweaver@bfalaw.com
adavis@bfalaw.com
mmontgomery@bfalaw.com
 Parties Represented:
 Bridgett Burk
 Cheryl Senko
 Jason Ariciu
 Jordan O'Hara
 Rev. Anthony Bell
 Samuel Armstrong
 Steven Akins
 Terry Fischer
 Tyler King

Matthew S. Melamed Esq.
Angelica M. Ornelas Esq.
Joshua D. Samra Esq.
Bleichmar Fonti & Auld LLP
555 12th St.
Suite 1600
Oakland, CA   94607-3616
Phone: 415-445-4003
mmelamed@bfalaw.com
aornelas@bfalaw.com
jsamra@bfalaw.com
 Parties Represented:
 Bridgett Burk
 Cheryl Senko
 Jason Ariciu
 Jordan O'Hara
 Rev. Anthony Bell
 Samuel Armstrong
 Steven Akins
 Terry Fischer
 Tyler King

Orin S. Snyder Esq.
Gibson Dunn & Crutcher
200 Park Ave.
47th Floor
New York, NY   10166-0193
Phone: 212-351-4000
osnyder@gibsondunn.com
 Parties Represented:
 Facebook, Inc.

Joshua S. Lipshutz Esq.
Gibson Dunn & Crutcher
1050 Connecticut Ave NW
Washington, DC   20036
Phone: 202-9558500
JLipshutz@gibsondunn.com
 Parties Represented:
 Facebook, Inc.

Russell H. Falconer Esq.
Gibson Dunn & Crutcher
2001 Ross Ave.
Suite 2100
Dallas, TX   75201
Phone: 214-698-3100
rfalconer@gibsondunn.com
 Parties Represented:
 Facebook, Inc.

Colin B. Davis Esq.
Gibson Dunn & Crutcher
3161 Michelson Dr.
Irvine, CA   92612-4412
Phone: 949-451-3800
cdavis@gibsondunn.com
 Parties Represented:
 Facebook, Inc.

Ms. Laura C. Mumm

Gibson Dunn & Crutcher
200 Park Ave.
New York, NY   10166
Phone: 212-351-4000
LMumm@gibsondunn.com
    Parties Represented:
    Facebook, Inc.

      I declare under penalty of perjury the foregoing to be true and correct. Executed at El Monte,

CALIFORNIA on November 8, 2021.

/s/ Anne Lieu
Anne Lieu
JAMS
alieu@jamsadr.com